Mass spectrometry-based proteomics of human cerebrospinal fluid

Biomarker discovery and verification in multiple sclerosis

Heidrun Vethe

This thesis is submitted in partial fulfillment of the requirements for the degree of Master of Science

Department of Biomedicine
University of Bergen

Bergen, Norway

June 2012

Acknowledgments

The presented work was carried out at the Multiple Sclerosis-group, PROBE, Department of Biomedicine, University of Bergen from August 2011 to June 2012. The work has been supported by The National Program for Research in Functional Genomics (FUGE) funded by the Norwegian Research Council, Helse Vest, the Meltzer Foundation, and the Kristian Gerhard Jebsen Foundation.

First, I would like to thank my supervisor Frode S. Berven. Thank you for always taking time to guide and for shearing your knowledge. You have been an excellent supervisor through out this whole process. A big thanks also goes to my co-supervisor Frode Selheim for helping me in handling the big gradient SDS-PAGE gel, thank you for your positivity and support in the writing process.

A special thanks goes to Ann Cathrine Kroksveen. Thank you for introducing me to SID-SRM analysis, for your motivating words and wonderful help both in the laboratory and in the writing process. To the rest of the hardworking people at PROBE, thank you, Elise and Yngvild for sharing your experiences as new master students in the field of proteomics. Thank you Thin Thin, for helping me with dimethyl-AIMS analysis. A big thanks also goes to Hilde and Anne for sharing your knowledge and lab-skills. Thank you Jill for your guidance and support in the writing process. To all the people at PROBE, I am very grateful for your time and support in this process. Thank you, Harald and Eystein for all your help, sharing your knowledge and computer skills. Thank you, Olav, you are the Orbitrap-master. Thank you, Astrid for being my “med-organisme” in this period.

Last but not least, a big thanks goes to my dear family and friends. Thank you for believing in me and motivating me throughout this whole process.

Bergen, June 2012
Heidrun Vethe

Summary

The main focus of this study was to use mass spectrometry-based proteomics to study protein abundance in cerebrospinal fluid (CSF) to reveal proteins that could serve as biomarker candidates in multiple sclerosis (MScl).

By combining a CSF pooling strategy and label-free relative quantification we discovered 65 proteins of differential abundance between MScl patients and controls. A selection of 17 biomarker candidates was further subjected to two independent verification steps: using stable isotope dimethyl labeling coupled to Accurate Inclusion Mass Screening (dimethyl-AIMS) and Stable Isotope Dilution Selected Reaction Monitoring (SID-SRM) for targeted quantification. The SID-SRM study included a larger patient cohort of 125 cases and controls. To our knowledge, this is the first report of a larger SRM verification study for biomarker candidates in MScl. The most interesting results from the biomarker discovery and verification study were the significantly decreased abundance of Apolipoprotein D, Cystatin C, Kallikrein-6 and Alpha-1-acid glycoprotein 1 in MScl patients compared to controls.

Furthermore, we performed a comprehensive characterization of the normal CSF proteome. We identified 18, 807 peptide mapping to 1987 proteins by applying immuno-affinity depletion and SDS-PAGE for enhanced proteome coverage. We obtain a comprehensive set of reference proteins that could further be used for investigations in MScl. The experiment gave us the opportunity to examine the size distribution on the SDS-PAGE gel of a selection of biomarker candidate proteins in normal CSF, with the aim to reveal potential protein variants (isoforms, truncation products and proteolytic processed products). We hypothesized that the identification of non-tryptic peptides could indicate truncation products of proteins in CSF. 13 and nine non-tryptic peptides were identified for the biomarker candidates Cystatin C and Secretogranin-1, respectively. This information had immediate utility for investigation in MScl. Based on the observed spread in size distribution of biomarker candidate proteins in normal CSF, we aimed to obtain quantitative information of proteins present in both high and low mass fractions on the gel and further target these proteins to obtain an abundance ratio between RRMS (patients) and OIND (controls), and investigate if these ratios differed for the same protein. The most striking observation was the opposite regulation level in CSF of Secretogranin-1.

Abbreviations

AIMS	Accurate inclusion mass screening
ACN	Acetonitrile
Ambic	Ammonium bicarbonate
BBB	Blood-brain barrier
BCSFB	Blood-CSF barrier
CID	Collision induced dissociation
CIS	Clinically isolated syndrome
CIS-MS	Multiple sclerosis patients that had CIS at lumbar puncture
CNS	Central nervous system
CSF	Cerebrospinal fluid
DTT	Dithiothreitol
ESI	Electrospray ionization
FA	Formic acid
HPLC	High pressure liquid chromatography
IgG	Immunoglobulin G
IAA	Iodoacetamide
LC	Liquid chromatography
MRI	Magnetic resonance imaging
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
m/z	Mass-to-charge ratio
MWCO	Molecular weight cut-off
MARS	Multiple affinity removal system
MScl	Multiple sclerosis
NOG	N-octyl-Beta-D-glucopyranoside
NN	Non-neurological
OCB	Oligoclonal bands
OIND	Other inflammatory neurological diseases
OND	Other neurological diseases
PAGE	Polyactylamide gel electrophoresis
PPMS	Primary progressive multiple sclerosis
SPMS	Secondary progressive multiple sclerosis
RRMS	Relapsing remitting multiple sclerosis
SDS	Sodium dodecyl sulphate
SRM	Selected reaction monitoring
TFA	Trifluoroacetic acid
Tris	Tris(hydroxymethyl)aminomethane
Q-Trap	Quadrupole-linear trap

Table of Contents

1 Introduction

1.1 Multiple sclerosis

Multiple Sclerosis (MScl) is an inflammatory and neurodegenerative disease associated with localized destruction of myelin sheaths and axons in the central nervous system (CNS) (Ferguson, Matyszak et al. 1997; Jacobs, Beck et al. 2000). The name of the disease refers to the multiple chronic sclerotic lesions scattered throughout the brain and spinal cord of MScl patients (Compston and Coles 2002), and could also refer to the multiple relapses and remissions characterized by the disease. A relapse is a symptom believed to result from lesion formation in the CNS whereof axonal demyelination suppresses the conduction of nervous signals (Noseworthy, Lucchinetti et al. 2000). Dependent on the location of the lesion, different symptoms occurs in the periphery of the CNS. Remission or recovery is a return to the baseline state that existed prior the last relapse and are believed to result from remyelination of the damaged myelin.

1.1.1 The multiple sclerosis patient

MScl is the most common disabling neurological disease affecting young adults, with disease onset normally occurring at the age between 20-40 (Rejdak, Jackson et al. 2010). A clear gender difference is observed whereof MScl affects twice as many women as it does men. Worldwide, the disease affects approximately 2.5 million people, predominantly affecting northern Europeans (Compston and Coles 2002). In Norway the prevalence of MScl is ~151-180/100 000 (Grytten, Glad et al. 2006). The overall cause of MScl is still unknown, the most accepted theory is that MScl is triggered by environmental factors, such as virus infections (Dalgleish 1997) or vitamin D deficiency, in genetic susceptible individuals (Sadovnick, Ebers et al. 1996).

1.1.2 Inflammation and neurodegeneration in multiple sclerosis

Myelin sheaths make up the principle target of immune attack in MScl (Figure I). Under normal conditions myelin sheaths serves as insulating segments needed for saltatory conduction, increasing the propagation efficiency of nervous impulses. As a consequence of the demyelinating events of MScl the signal transduction along the axons in the CNS is impaired, thus hampering the transmission of signals from CNS to the periphery. The area in the periphery may become affected with a variety of symptoms. The signs and symptoms of relapses are diverse, and include among others optic neuritis, weakness in limbs, cognitive impairment, tremor and fatigue (Compston and Coles 2008).

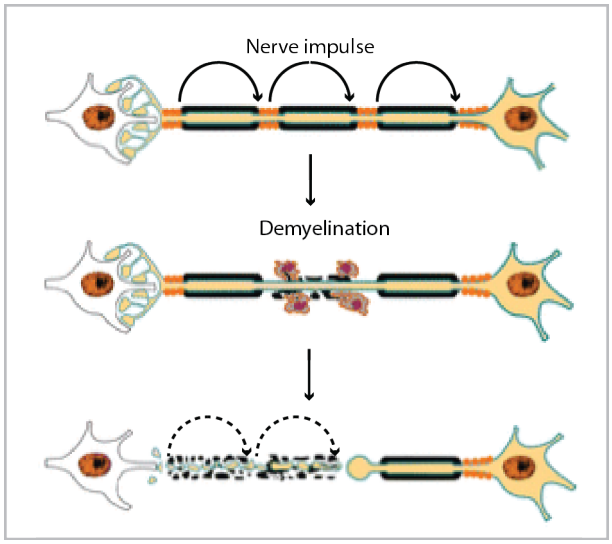

Figure I: Demyelination. A normal axon contains intact myelin sheaths that increase the efficiency of action potential propagation along the nerve cell. Due to the demyelinating process of MScl axons in lesion areas destroyed and the conduction of axonal impulses are inhibited. Adapted from (Trapp and Nave 2008).

The hallmarks of multiple sclerosis

The pathology of MScl is distinguished from that of other inflammatory diseases of the CNS by the presence of chronic multifocal sclerotic lesions scattered throughout the brain and the spinal cord (Lassmann 1998). These lesions represent the hallmark and endpoint of MScl pathology, and consist of localized areas characterized by ongoing demyelination; resulting in loss of myelin, relative axonal destruction and glial scar formation (gliosis) (Noseworthy, Lucchinetti et al. 2000).

Development of lesions – proposed disease mechanisms

The evolution of MScl lesions involves several steps: immune engagement, demyelination, acute inflammatory injury of axons and glial, limited remyelination, gliosis and neurodegeneration (Compston and Coles 2002; Bielekova and Martin 2004). MScl is believed to be of autoimmune origin and initiated by activation of auto-reactive T cells against components of the oligodendrocyte-myelin unit. These T cells cross the blood-brain barrier (BBB) and gain access to the CNS (Trapp, Bo et al. 1999; Bruck and Stadelmann 2003). Once inside the CNS, leukocytes and macrophages along with local glial cells, target specific areas of neurons and destroys myelin, causing relative axonal destruction and loss. At disease onset, inflammation is transient and remyelination occurs, however over time the pathological changes become dominated and eventually result in chronic neurodegeneration (Compston and Coles 2002).

1.1.3 Disease heterogeneity and clinical presentation

MScl is a complex disease, the combination of unpredictable clinical course and variations in mechanisms of myelin destruction across patients contribute to the heterogeneity of the disease (Lucchinetti, Bruck et al. 2000; Bruck and Stadelmann 2003). Early events in the disease pathway of MScl are often clinically silent and may not result in any acute signs of the disease (Compston and Coles 2002), therefore neurological damage may be present before the first symptoms occur. The first indication of MScl is termed clinically isolated syndrome (CIS) of demyelination (Miller, Barkhof et al. 2005). Normally full recovery from the first episode of symptoms is observed. Over time, however, recovery from each relapse is incomplete and persistent symptoms accumulate. 60-80% of CIS patients convert to clinical definite MScl in the follow-up period (Miller, Chard et al. 2012). CIS and progression to MScl is unpredictable. The majority of MScl patients (85-90 %) have the relapsing remitting (RRMS) subtype, characterized by clearly defined relapses divided by periods of full recovery (Figure II). After a period of RRMS (normally 15-20 years) most patients progress into a secondary progressive (SPMS) disease course, in which the disease progressively worsens (Keegan and Noseworthy 2002). In the primary-progressive (PPMS) type (affecting 10-15%), the disease is progressive in nature at

onset and disability increases over time without periods of improvements (Noseworthy, Lucchinetti et al. 2000).

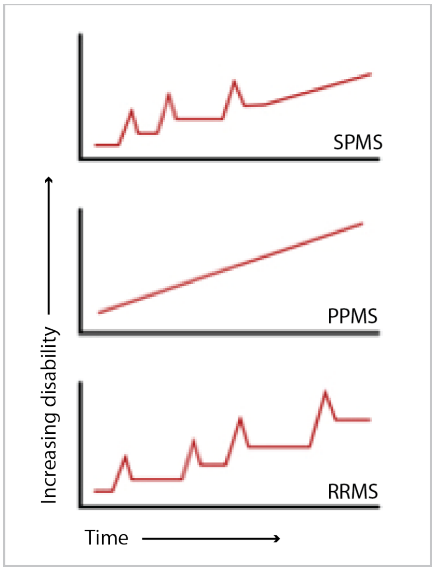

Figure II: Clinical subtypes of MScl. Based on the disease course of MScl, three main subtypes are recognized. A relapsing remitting subtype, affecting 80-85% of all MScl patients, that further could progress to secondary-progressive MScl. 10-20% of MScl patients have a progressive disease course from onset (PPMS).

The median life expectancy of a MScl patient is around 30 years from disease onset (Compston and Coles 2008). Although there is no cure for MScl, early detection and subsequently early treatment will retard the long-term evolution of the disease. Treatment aims to slow down the disease progression, prevent axonal damage as well as reduce the frequency of relapses. The efficiency of treatment varies depending on the subgroup as well as progression of the disease (Compston and Coles 2008).

1.1.4 Current diagnosis of multiple sclerosis

Current diagnosis of MScl are mainly based on the McDonald (McDonald, Compston et al. 2001) and revised McDonald Criteria (Polman, Reingold et al. 2005), and relies in addition to disease history, on clinical examination supported by laboratory investigations. Magnetic resonance imaging (MRI) is performed to visualize lesions, detecting the demyelinating side of the disease. Furthermore, CSF biochemistry measurements that include assessment of oligoclonal immunoglobulin G (IgG) bands are performed in order to detect the inflammatory side. The presence of oligoclonal IgG bands in CSF and further absence in blood serum, imply that immunoglobulin is produced in the CNS. MScl are considered clinically definite when neurological

dysfunction becomes “disseminated in time and space” (Noseworthy, Lucchinetti et al. 2000). The presented diagnostic findings are not disease specific for MScl, but could also indicate other inflammatory neurological diseases (OIND), which emphasizes the urgency of developing novel protein biomarkers that are exclusive for MScl.

1.2 Definition of biomarkers

The National Institutes of Health (NIH) officially defines a biomarker as *"a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention."* In the presented study a biomarker refers to a protein measured in a body fluid whose concentration reflects the presence of MScl. Of particular interest are proteins with relevance to biochemical and physiological changes involved in the transformation from normal to diseased state. Due to the complexity and heterogeneity of MScl, there is probably no single protein or peptide that could serve as a biomarker for the disease in a clinically relevant way. It is doubtful that one single biomarker could reflect all ongoing pathological processes in addition to predict disease progression in all the disease subtypes. However, there is growing consensus that a set of disease specific proteins could be used in a biomarker panel (Ottervald, Franzen et al. 2010). Such a biomarker panel could be useful in the clinic to diagnose, monitor disease activity and predict disease progression (Rifai, Gillette et al. 2006).

1.2.1 Biomarker discovery in multiple sclerosis

Biomarker discovery strives to discover proteins that are either present, absent or of differential abundance in a large percentage of disease subjects and not in most subjects without the disease. In MScl, biomarker discovery is often based on studying protein abundance levels in CSF with the aim to identify proteins that differentiate between MScl and controls and that can potentially be used to diagnose, monitor clinical course, progression and treatment effects. In addition the biomarkers could add biological information that may provide insight into disease mechanisms of MScl.

Biomarker discovery can generally be performed in a wide range of biological material, such as body fluids and disease-affected tissue. In relations to MScl, lesions are of special interest as they are the primary location of disease specific activity and hallmarks of the disease. Unfortunately, due to inaccessible location, lesions are only biopsied post-mortem and could therefore be affected by processes that are not disease specific. For biomarker studies in living subjects, the body fluid most proximal to the MScl lesions represents a suitable choice. Cerebrospinal fluid (CSF) surrounds the CNS and fills the brain, the spinal cord as well as the subarachnoid space (Segal 1993). Thus CSF is in direct contact with the CNS and reflects in this way the biochemical state of the CNS under different physiological and pathological settings. Therefore, CSF is expected to reflect disease related alterations of its surrounding tissue (Harris and Sadiq 2009). Approximately 80% of the protein composition in CSF resides form blood, while the remaining 20% is derived directly from the CNS (Segal 1993; Tumani, Hartung et al. 2009). In this manner CSF is rich in brain specific proteins as well as it reflects the composition of blood plasma, although the concentration of proteins in CSF is lower. The crude protein concentration in CSF ranges from 0.2 to 0.8 mg/mL (0.3-1% of serum protein concentration) (Regeniter, Kuhle et al. 2009), though the concentration may vary due to individual variations such as overall health of the patient and biological variations concerning among others the Circadian rhythm (Nilsson, Stahlberg et al. 1992) and CSF flow rate (Reiber 1994). The CSF volume in a healthy human adult is approximately 150 mL (Smith, Johanson et al. 2004), and around 500 mL of CSF is produced each day. A typical volume of 10-15 mL are collected by lumbar puncture for laboratory investigations (Blennow, Fredman et al. 1993).

The blood-brain barrier (BBB) and the blood-CSF barrier (BCSFB) separate CSF from the bloodstream (Smith, Johanson et al. 2004), and acts as gates for the exchange of proteins between the two body fluids. Thus, the protein composition of CSF is a combination of blood derived proteins as well as proteins produced locally by the CNS. Disease mechanism of MScl and other inflammatory neurological diseases are proposed to alter the protein composition of CSF, whereof BBB dysfunction could result in increased abundance of blood-derived proteins in CSF (Reiber 2001). The diffusion across the two barriers is under normal conditions size-

dependent (Smith, Johanson et al. 2004). Larger blood derived proteins, like Apolipoprotein B and Hemoglobin are excluded and normally not present in CSF, hence these proteins could be used to check CSF for blood contamination (Zhang 2007). Blood contamination of CSF may arise during lumbar puncture and even a small percentage of blood could dramatically alter the CSF proteome when analyzed (You, Gelfanova et al. 2005).

1.2.2 The Biomarker pipeline

Development of biomarkers is divided into four general phases, constituting of the discovery phase, the qualification phase, the verification phase and the validation phase (Rifai, Gillette et al. 2006). Biomarker discovery in MScl is performed in a relatively small number of well-characterized CSF samples. By comparing the proteomes of cases and controls one can detect differentially abundant proteins between the groups. In order to confirm the differential expression level, the candidate biomarkers are subjected to several additional quantification steps. The qualification phase is the first line of confirmation, in which biomarker candidates derived from the discovery phase are targeted for quantification in individual samples by use of alternative analytical methods. Verification of biomarker candidates is performed in larger patient cohorts by incorporating a boarder range of cases and controls. Biomarker validation is performed on the few candidate biomarkers that performed well in the verification phase, in which clinical assays like enzyme-linked immunosorbent assay (ELISA) are usually established. Finally, validated biomarkers may be selected for 'commercialization', in which research immunoassays are adapted for clinical testing (Rifai, Gillette et al. 2006).

1.3 Proteomics

The term proteome is a combination of "PROTEin" and "genOME" and refers to the total protein content expressed by a genome at a given time (Wilkins, Pasquali et al. 1996). The proteome is the dynamic product of the genome, and due to various layers of protein processing; the proteome is prone to change in relations to different conditions. The proteome is the entire complement of proteins and their variants

produced by an organism, such as the human proteome, or of a limited system, such as the CSF proteome. Proteomics is the large-scale study of the proteome (Mallick and Kuster 2010), and involve technologies for identification and quantification of a large portion of the protein content, enabling the study of the complex and dynamic nature of proteins. Hence, proteomics analyses are performed in order to characterize proteomes as well as identify and quantify proteins of differential abundance by comparing distinct proteomes. For this purpose, mass spectrometry (MS) is a versatile technology available to analyze endogenous proteins in complex biological samples (Aebersold and Mann 2003).

1.3.1 Mass spectrometry-based proteomics

The basic principle of mass spectrometry (MS) involves a technology that enables the measurement of mass-to-charge (*m/z*) ratio of ions in gas phase to determine the masses of peptides (Han, Aslanian et al. 2008). A mass spectrometer consists of an *ion source*, performing the ionization of peptides, a *mass analyzer* that separates the ions according to their *m/z* ratio and a *detector* that records the number of peptide ions at each *m/z* value (Aebersold and Mann 2003).

Prior to being subjected to MS analysis, the peptides can be subjected to liquid chromatographic (LC) separation that is directly connected to the MS (LC-MS). The resulting chromatogram displays the eluting peptides as peaks with a given intensity and retention time.

1.3.2 Protein identification by use of mass spectrometry

Proteins are identified based on their corresponding peptides, either by the accurate mass or by the amino acid sequence of the peptides. Peptide mass fingerprinting (PMF) involves protein identification by searching the peptide masses measured by mass spectrometry against a database containing theoretical peptide masses (Gevaert and Vandekerckhove 2000). In order to obtain more detailed and reliable identification, tandem MS is applied. In tandem MS the peptide are further fragmented into its respective amino acids.

1.3.3 Tandem mass spectrometry

In tandem MS the mass spectrometer measures the masses of all the peptides eluting at any given time from the chromatographic column, and subsequently select a number of peptide ions with the highest intensities for fragmentation. This is performed by allowing only ions of a particular mass through a collision cell, where they are provided with sufficient energy for the peptide bonds to break (de Hoog and Mann 2004). The resulting spectrum is called tandem mass spectrum (MS/MS). The tandem spectrum contains several adjacent fragments that spell out the amino acid sequence of the peptide in question (de Hoog and Mann 2004). The fragment masses are further compared to theoretical fragment masses calculated from all the peptides in protein databases to enable protein identification. Typical databases that can be used are SwissProt and UniProt, in which UniProt is unreviewed and automatically annotated, while the SwissProt database is reviewed and manually annotated with information extracted from the literature. There are several different search engines available, with Mascot being the most widely used. The in house developed software SearchGUI (Vaudel, Barsnes et al. 2011) combine the results from two search engines simultaneously, namely the Open Mass Spectrometry Search Algorithm (OMSSA) and X!Tandem. For visualization of the peptide and protein identifications, the in house developed software Peptide Shaker can be applied.

1.3.4 Sample processing prior mass spectrometry

Proteins of low abundance are of special interest in biomarker discovery, as these proteins are likely to be present as a result of ongoing disease specific processes. Detection of proteins of low concentration may be troublesome using MS, as they are often masked by the high concentration of more abundant proteins. In order to get broader proteome coverage, the proteome of a complex biological sample needs to be reduced. Fractionation strategies can be applied both on protein and peptide level. Separation on protein level includes separation using molecular weight (MW) cut-off filters, targeted protein depletion and separation by use of gel based techniques. Sample processing typically begins with depleting the sample of high abundant proteins. For this purpose several immuno-affinity depletion column has been introduced with antibodies targeting a selection of proteins. SDS-PAGE is among the

most used methods separating proteins according to size; hence enable identification and evaluation of proteins variants (isoforms, truncation and proteolytic products). The term gel based liquid chromatography mass spectrometry (GeLC-MS) combines the separation of proteins using SDS-PAGE followed by gel cutting, protein digestion, online reverse phase nano-LC separation, and MS analysis of the peptides from each gel band (Schirle, Heurtier et al. 2003). Before the protein mixture can be subjected to MS analysis, proteins are enzymatically digested into peptides. Trypsin is the most used protease for this purpose, and hydrolyzes peptide bonds at the carboxyl terminal side of lysine and arginine residues. On the peptide level, further fractionation can be applied, such as ion exchange chromatography (Strong Cation Exchange (SCX) and Strong Anion Exhange (SAX)). The term multidimensional protein identification technology (MudPIT) describes the use of multidimensional liquid chromatography prior MS analysis for protein identification (Washburn, Wolters et al. 2001). In general, the more separation and fractionation methods combined, the better coverage of the proteome is expected.

1.4 Quantitative proteomics

Quantitative proteomics aims to obtain quantitative information of the abundance of proteins and peptides in complex biological samples. Furthermore, this branch of proteomics aims to compare two or more distinct proteomes to identify proteins with altered abundance levels between the comparison partners (Bantscheff, Schirle et al. 2007). Mass spectrometry-based proteomics enable both relative and absolute quantification of peptides and proteins (Aebersold and Mann 2003). Quantitative information can be obtained using non-targeted and targeted strategies.

1.4.1 Non-targeted quantification strategies

Biomarker discovery studies aim to identify and quantify proteins that are of differential abundance between two or more physiological states. Discovery of such biomarker candidates are possible by comparing different proteomes. Relative quantitation, determination of a ratio between a proteins concentration in one sample versus that of another, can include labeling of the compared groups or a label-free

approach. Labeling strategies involves chemical incorporation of stable isotopes to label peptides in a sample, and are based on the assumption that a stable isotope labeled peptide have chemically identical properties to its endogenous counterpart, thus the peptides will behave identically during LC and MS analysis (Bantscheff, Schirle et al. 2007). Because the labeling reagent introduce a mass shift, the same peptide can be quantified separately in the same MS analysis, and their ratio represent the relative abundance of the corresponding peptides are obtained (Nesvizhskii, Vitek et al. 2007). Chemical incorporation of stable isotopes includes among others dimethyl labeling and iTRAQ. Stable isotope dimethyl labeling involve the use of formaldehyde to globally label the N-terminus and the amino group on lysine through reductive amination (Hsu, Huang et al. 2003). For relative quantification of stable isotope labeled peptides, MS and MS/MS information are obtained. The MS/MS information is used for protein identification, while the MS information is used for quantification. The software MaxQuant and Proteome Discoverer Daemon are commonly used for quantitative proteomics analysis.

Label-free quantification uses no labeling of the different proteomes to be compared, instead comparison is performed by aligning individual sample LC-MS runs. The approach involves comparison of chromatographic and spectral analysis and is therefore dependent on similar conditions during LC and MS analysis of the samples. Strategies involve both measuring and comparing the mass spectrometric signal intensity of peptide precursor ions belonging to a specific protein (intensity based) (Mallick and Kuster 2010). Counting and comparing the number of fragment spectra, identifying peptides of a given protein (spectral count) (Mallick and Kuster 2010) or peptide abundance can be determined from the intensity of corresponding spectrum features identified by the mass spectrometer (feature based) (Nesvizhskii, Vitek et al. 2007). The software MaxQuant, Spectrum Mill and Progenesis LC-MS are used for label-free relative quantification.

1.4.2 Targeted quantification

Accurate Inclusion Mass Screening

Accurate Inclusion Mass Screening (AIMS) is a targeted MS method for detection of selected peptides derived from target proteins. Masses and charge state of unique peptides representing target proteins are incorporated into an including list. The MS is instructed to monitor the included peptides in each MS scan. MS/MS spectra for sequence confirmation are acquired when a peptide form the list is detected with both the correct mass and charge state (Jaffe, Keshishian et al. 2008).

Selected reaction monitoring

Selected reaction monitoring (SRM) is a targeted MS method that enables quantification of a specific subset of proteins over a large number of samples, using peptides as surrogates for the target proteins (Gallien, Duriez et al. 2011).

SRM is performed on a triple quadrupole (QQQ) mass spectrometer. In the triple quadrupole instrument two steps of mass filtering is performed in combination with collision-induced dissociation (CID). The instrument can be instructed to select and isolate the *m/z* value of the precursor peptide (peptide ion) in the first quadrupole (Q1). The peptide ion then enters the collision cell (Q2) and is fragmented. Predefined fragments (fragment ion) of the precursor peptide (fragment ion) are further monitored in the third quadrupole (Q3). The selected precursor/fragment ion pairs are referred to as transitions and are *m/z* settings for Q1 and Q3 (Lange, Picotti et al. 2008). Thus, in SRM, the selected precursor to fragment ion "reaction" is monitored. The SRM methodology provides relative and absolute quantification of protein and peptide when appropriate standards are added.

Stable isotope dilution (SID) in combination with SRM is the gold standard for absolute quantification (Gallien, Duriez et al. 2011). The absolute quantification (AQUA) approach (Kirkpatrick, Gerber et al. 2005) allows for both relative and absolute quantification. Absolute quantification (AQUA) of endogenous peptides requires comparison with internal standards of known concentrations. Stable isotope labeled internal standards (SISs) is synthetic homologs to the endogenous peptide

enriched with stable isotopes. Thus SISs have a higher mass and are referred to as 'heavy' peptides, as appose to their endogenous counterpart recognized as 'light' peptides. In SID-SRM analysis, light and heavy transitions are analyzed for every precursor and fragment combination (Lange, Picotti et al. 2008). Dependent on if the concentration of the SIS is precisely determined; the absolute or relative concentration of the target peptide can be determined from the relative intensity of the light/heavy transitions.

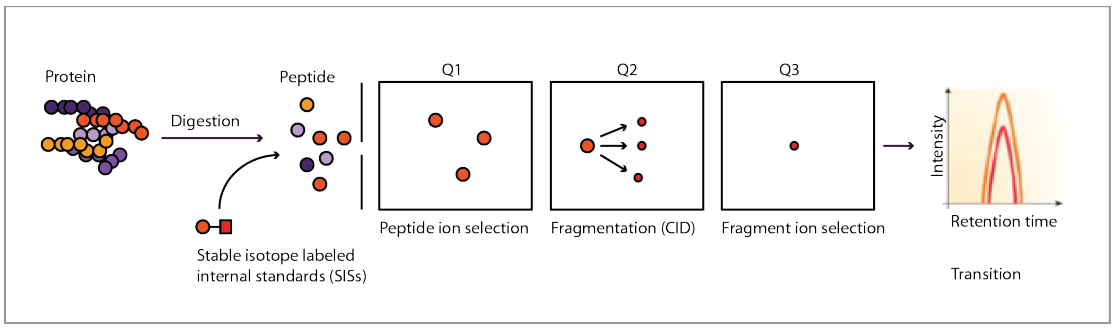

Figure III: Targeted protein quantification using stable isotope internal standards (SISs) and selected reaction monitoring (SRM). Targeted protein quantification is achieved by SRM on a triple quadrupole mass spectrometer (MS). The protein sample is digested and the peptide mixture is spiked in with stable isotope labeled internal standards (SISs). The peptide mixture is fractionated and ionized by liquid chromatography (LC) and electrospray ionization (ESI). The MS are instructed to target a list of peptides and fragments thereof, corresponding to the target proteins. In the first quadrupole (Q1) a specific peptide ion is selected. The peptide ion is fragmented by collision-induced dissociation (CID) in the second quadrupole (Q2). The corresponding fragment ion is filtered in the third quadrupole (Q3). The resulting peptide/fragment ion pairs are referred to as a transition and are highly specific for the given peptide.

SRM data can be analyzed using the software Skyline or the software Analyst® and MultiQuant™, in which SISs and endogenous peptides are integrated and the most abundant transitions free of interference are used for quantification. The median SRM area ratio is used to quantify the abundance difference of proteins between two states.

Aims of the study

The overall aim of the presented study was to use mass spectrometry-based proteomics to study the human cerebrospinal fluid (CSF) proteome in search for biomarker candidates in multiple sclerosis. The presented study was divided into five different experiments.

1. **Biomarker discovery in multiple sclerosis**

The aim of the biomarker discovery experiment was to compare the neat and depleted CSF proteome of multiple sclerosis patients with controls by applying a CSF pooling strategy and label-free relative quantification of protein of differential abundance between the patient groups.

2. **Qualification of biomarker candidates using accurate inclusion mass screening**

The aim of the qualification experiment was to apply stable isotope dimethyl labeling and Accurate Inclusion Mass Screening (dimethyl-AIMS) for targeted quantification of the biomarker candidates in novel and individual CSF samples.

3. **Verification of biomarker candidates using selected reaction monitoring**

The aim of the verification experiment was use Stable Isotope Dilution Selected Reaction Monitoring (SID-SRM) for targeted quantification of discovered and literature-derived biomarker candidates in a larger patient and control cohort.

4. **Characterization of the normal human cerebrospinal fluid proteome**

The aim of the experiment was to do a comprehensive mapping of the CSF proteome, and to obtain an overview of the molecular weight distribution of certain biomarker candidates in CSF.

5. **Evaluation of size dependent differential abundance of CSF protein variants in multiple sclerosis**

The aim of the experiment was to obtain an abundance ratio between multiple sclerosis and controls for proteins present in both high and low molecular mass portions of the CSF proteome and to see if the two ratios of the same protein differed in multiple sclerosis.

2 Materials and methods

All materials used in the following methods are listed in Table 2.1.1 and 2.1.2. Unless stated otherwise, all additional chemicals were purchased from commercial sources and were of analytical grade.

Table 2.1.1 Chemicals

Supplier	Chemical	Quality	Catalog number
Agilent Technologies	Depletion Buffer A		
	Depletion Buffer B		
Amersham Biosciences	Coomassie Brilliant Blue tablet		
	DiThioTreitol (DTT)		
	Tetramethylethylenediamine (TEMED)		
Biorad	30% Acrylamide/ 2.67% Bis		
	Ammonium persulphate (APS)		
	10% SDS solution		
Fluka Analytical	Acetonitrile (ACN) LC-MS Chromosolv®	MS grade	
	Acetic acid (HAc)		
	Formic acid (FA)	MS grade	
	Sodium cyanobromohydride ($NaBH_3CN$)		
Invitrogen, Life Technologies	SeeBlue® Plus2 Pre-Staining Standard (1x)		
	LDS Sample Buffer (4x) NuPage®		
	4-12% Bis-Tris Gel NuPage®		
	MES SDS Running Buffer (20x) NuPage®		
Isotec™	Formaldehyde-^{13}C-d_2 ($C_{13}D_2O$)		
	Formaldehyde-h_2 (CH_2O)		
	Formaldehyd-d_2 (CD_2O)		
Merck	Glysine		
Promega	Trypsin Porcine		V511A
	Trypsin Resuspension Buffer (acetic acid)		V542A
Sigma-Aldrich	Urea Sigma Ultra		
	Methylamine 40 wt %		
	Iodoacetamide (IAA)		
	Trifluoroacetic acid (TFA)		
	Tris(hydroxymethyl)aminomethane		
	Calcium chloride ($CaCl_2$) x 2 H_2O		
	Ammonium bicarbonate		
	Water Chromosolv® Plus for HPLC	MS grade	
	Metanol Chromosolv® (MeOH)		
	Triethylammonium biacrbonate (TEAB)		
	Sodium cyanoborodeuteride ($NaBD_3CN$)		
	N-octyl-Beta-D-glucopyranoside (NOG)		

Table 2.1.2 Materials

Supplier	Material
Agilent Technologies	Spin filter, 0.22 µm
Millipore	Ultra-4 Centrifugal Filter Unit (MWCO filters)
Waters	Oasis® HLB µElution plate 30 µm
3M Empore™	C18 StageTip
	C8 StageTip

Table 2.1.3 Kit

Supplier	Kit	Content
Invitrogen, Life Technologies	Qubit™ Quantification Kit	Quant-it™ protein reagent
		Quant-it™ protein standard 1
		Quant-it™ protein standard 2
		Quant-it™ protein standard 3

Table 2.1.4 Apparatus

Supplier	Apparatus
Amersham Biosciences	Electrophoresis Power Supply (EPS 601)
	Gel-casting Pump
Biorad	PROTEAN II xi vertical electrophoresis cells
	Gradient Former model 285
	Gel casing plates and additional equipment
Eppendorf	Concentrator 5301
	Centrifuge 5810R
	Thermomixer Comfort (Eppendorfmixer)
Invitrogen, Life Technologies	Novex Mini-Cell
Heraeus	Biofuge Statos

Table 2.1.5 Mass Spectrometers

Supplier	Mass spectrometer
AB SCIEX	Q-Trap® 4000 LC-MS/MS Systems
AB SCIEX	Q-Trap® 5500 LC-MS/MS Systems
Thermo Scientific	Orbitrap Velos Pro

Table 2.1.6 Liquid Chromatography system

Supplier		
Dionex Ultimate	3000RS nano-LC system	Pre-column (Dionex, Acclaim PepMap Nano Trap column, C18, 75µm i.d. x 2 cm, 3 µm) Analytical column (Dionex, Acclaim PepMap 100 RSLCnano column, 75µm x 15cm, C18, 2 µm)

Table 2.1.7 Software

Supplier	Software
AB SCIEX	Analyst® 1.5.1
AB SCIEX	MultiQuant™ 2.1.2
Dionex Corporation	Chromeleon™
Agilent Technologies	Spectrum Mill
In house developed	SearchGUI
In house developed	Peptide Shaker
Thermo Scientific	Proteome Discoverer
Thermo Scientific	Proteome Discoverer Daemon

Stable isotope labeled internal standards (SISs) peptides used is listed in Supplementary Table 6.6 and 6.7. *Due to the size and format an Additional appendix (Supplementary Tables 3.1B-3.5B) is available at a local computer (EIR server) at PROBE.*

2.1 Clinical samples

The CSF samples used for CSF proteome characterization and biomarker discovery in multiple sclerosis (MScl) were obtained from four different locations: (I) Department of Neurology, Haukeland University Hospital, Bergen, Norway, (II) Neurology Department UCL, Université Catholique de Louvain, Brussels, Belgium, (III) Laboratory of Neuroimmunology, IRCCS, "C. Mondion Neurological Institute", Pavia, Italy. All locations are part of the European BioMS-eu consortium for CSF biomarker research in multiple sclerosis (www.bioms.eu). Neurological normal CSF samples were collected from patients receiving spinal anesthesia prior to non-neurological minor surgical investigations at the (IV) Department of Orthopedic surgery, Haukeland University Hospital, Bergen, Norway. These subjects were diagnosed as having no neurological diseases and were considered to have neurologically normal CSF.

All CSF samples were collected according to the published consensus protocol for CSF collection and biobanking (Teunissen, Petzold et al. 2009). In brief, immediately after the CSF was collected by lumbar puncture, the CSF samples were centrifuged at 450 *x g* for 5 minutes to discard cells and cell debris, and the supernatant was stored at -80 ^{o}C, until further sample processing. The protein concentrations in crude CSF

were measured using a Qubit™ fluorometer (Invitrogen, Carlsbad, CA, USA) according to the instructions given by the vendor.

2.1.1 Patient selection

The multiple sclerosis (MScl) patients included in this study were initially diagnosed according to the revised criteria of McDonald (Polman, Reingold et al. 2005). Patients and controls were divided into six categories: (I) CIS: patients with CIS that had not developed clinically definite (CD) multiple sclerosis in the follow-up period. (II) CIS-MS: patients with CD multiple sclerosis diagnosed as CIS at the time of lumbar puncture, (III) RRMS, (IV) controls with other neurological diseases (OND), (V) controls with other inflammatory neurological diseases (OIND), and (VI) spinal anesthesia subjects diagnosed as non-neurological (NN). All patient information is available in Supplementary Table 6.1-6.5.

2.2 Processing of CSF samples - protein separation strategies

2.2.1 Concentration and purification of CSF samples

Ultra-4 Centrifugal Filter Units (MWCO filters, Millipore, Billerica, MA, USA) containing Ultracel-3 membrane were used to desalt and concentrate CSF samples prior further processing. Salts and proteins less than 3 kDa was filtered out and thereby be excluded form the samples.

The filters were washed with deionized water (Milli-Q water, Millipore, Billerica, MA, USA) and centrifuged at 3000 *x g* for 5 min at 4 °C (Centrifuge 5810R, Eppendorf AG, Hamburg, Germany). The total volume of CSF sample were added to the filters and centrifuged at 3000 *x g* for 30 min at 4 °C followed by an additional centrifugation round with 1 mL deionized water, and subsequent centrifugation until about the volume of 100 µL remained in the filter. Finally, the samples were concentrated to dryness by using a vacuum concentrator (Concentrator 5301 from Eppendorf AG, Hamburg, Germany).

Current MS-based proteomics is often biased towards abundant proteins, which may mask the detection of less abundant proteins. Therefore, the complexity of CSF proteomes was reduced prior to MS analysis. CSF samples were separated on protein level by performing two different strategies; immuno-affinity depletion of 14 abundant proteins in CSF and SDS-PAGE strategies that separates proteins according to molecular size.

2.2.2 Immuno-affinity depletion

A human Multiple Affinity Removal System (MARS HU-14) 4.6 mm x 50 mm LC column (Agilent Technologies, Santa Clara, CA, USA) was used for immuno-affinity depletion of 14 abundant proteins (albumin, IgG, antitrypsin, IgA, IgM, transferrin, haptoglobin, alpha-1-acid glycoprotein, alpha-2-macroglobulin, apolipoprotein A-I and A-II, fibrinogen, C3 and apoliprotein B). According to the vendor, depletion of these 14 proteins accounts for approximately 94 % of the total protein amount.

Dried and concentrated CSF samples were dissolved in 200 µL MARS Buffer A and further filtered through a 0.22 µm CTA filter (both from Agilent Technologies, Santa Clara, CA, USA) on 2000 *x g* at room temperature for 5 min, followed by injection onto the MARS column. The MARS column was connected to a Dionex Ultimate 3000 nano-LC system (Thermo Scientific, Waltham, MA, USA) was equilibrated with MARS Buffer A.

The gradient for the MARS column with dimensions 4.6 mm x 50 mm was as follows; for the first 9.5 minutes, 100 % Buffer A with a flow rate of 125 µL/min was used and the flow-through with depleted CSF proteins were collected between 3.2-13 minutes. The bound fraction was collected from 15.5-21 minutes with 100 % MARS Buffer B with the flow of 1000 µL/min. The column was furthermore run in 100% MARS Buffer A to regenerate the column before the next sample was injected.

The flow-through and protein depleted CSF was purified and concentrated with 3 kDa MWCO filters as described above, pre-rinsed with 0,1 % N-octyl-β-Dglycopyranoside (NOG) (Sigma-Aldrich, St.Louis, MO, USA) for increased protein recovery. Samples

were subsequently dried and concentrated at 30 °C by use of a centrifugal vacuum concentrator.

2.2.3 Separation of proteins using 5-15% gradient SDS-PAGE

Gel casting

In order to separate proteins according to size, a 20 cm 5-15% sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE) gel was casted. The gel casting solutions were as follows; 5% acrylamide gel (28,5 mL Milli-Q water, 12,5 mL 1.5 M Tris-Hcl (pH 8.8), 8.4 mL 30% Acrylamide/ 2.67 % Bis and 500 µL 10% SDS). The 15% acrylamide solution contained the same amount of 10% SDS and 5.6 g Sucrose, in addition to 8.7 mL Milli-Q water, 12.5 mL 1.5 M Tris-HCl (pH 8.8), 25 mL 30% Acrylamide /2.67 % Bis. The 5% and 15% solutions were made and transferred to a gradient former (Biorad, Hercules, CA, USA), followed by direct addition of 20 µL 10% ammonium persulfate (APS) and 8 µL Tetramethylethylenediamine (TEMED) (Amersham Biosciences, Piscataway, NJ, USA). The gradient solutions were added to the gel casing plates and left until polymerized. Thereafter, the stacking gel was casted, by mixing 12.2 mL MilliQ water, 5.0mL 0.5M Tris-HCl (pH 6.8), 2.6 mL 30% Acrylamide / 2.67 % Bis and 200 µL 10 % SDS followed by addition of 100µL 10 % APS and 20 µL TEMED to induce the polymerization.

Loading and running SDS-PAGE gel

CSF samples were reduced and alkylated prior to loading onto the gel. For reducing the CSF samples; 4 x LDS Sample Buffer (Invitrogen, Carlsbad, CA, USA) was diluted in Milli-Q water to 2 x, the reduction agent DTT (Amersham Biosciences, Piscataway, NJ, USA) was added to give a final concentration of 10 mM. The resulting sample solution was heated for 5 minutes at 100 °C and further cooled in room temperature before addition of the alkylating agent, iodoacetamide IAA (Sigma-Aldrich, St. Louis, MO, USA). IAA was added to give a final concentration of 20 mM, followed by incubation in the dark for 20 minutes. Samples were loaded and the run at 60 V for 16 hours in 1 x electrode running buffer (25mM Tris base, 192 mM Glysine, 0.1% SDS diluted in milliQ water).

Gel staining

After protein separation, the gel was stained with Coomassie Brilliant Blue (Amersham Biosciences, Piscataway, NJ, USA) in 30% methanol MeOH and 10% acetic acid for 2 hours and distained in 30% ethanol/ 10% acetic acid over night.

2.2.4 Separation of proteins using 4-12% NuPage® Bis-Tris gel

For separation of proteins according to size over a short time period a 4-12 % Bis-Tris Gel, NuPage® (Invitrogen, Carlsbad, CA, USA) (1.0mm x 10 wells) was used for SDS-PAGE. Reduced and alkylated CSF protein samples were loaded onto the wells and the gel was run at 200V for 20 minutes in MES-buffer (Invitrogen, Carlsbad, CA, USA). The gel was stained using Coomassie Brilliant Blue and distained in 30% ethanol/ 10% acetic acid as described above.

2.3 Digestion protocols

2.3.1 In-solution protein digestion using trypsin

The dried CSF sample was dissolved and gently mixed in 20 μL urea solution (480 mg Urea, 1.7 μL 40 wt% methylamine in H_2O and 630 μL dH_2O). 20 μL trypsin buffer (0.61 g Tris and 15 mg $CaCl_2$ x H_2O) (all from Sigma-Aldrich) was added to each of the samples followed by incubation at room temperature with gentle shaking (300 rpm) in an Eppendorf Mixer. Each of the subsequent steps involved incubation periods of one hour (in the dark for IAA) after addition of 4 μL 100 mM DTT and 5 μL 200 mM IAA (18.5 mg IAA in 500 μL dH_2O) respectively. To avoid unwanted alkylation, 0,8 μL 100 mM DTT was added and samples where further incubated for 10 minutes at room temperature. The samples were digested with Trypsin Porcine (Promega, Fitchburg, MO, USA) at a 1:50 trypsin to protein ratio, and incubated for 15-17 hours (over night) in shaker at 37 °C with gentle shaking as above. To stop the reaction, 15 μL 10% Formic Acid (FA) (Fluka Analytical, Buchs, Switzerland) was added to each sample prior concentrating the CSF samples to dryness by use of a vacuum concentrator.

In-solution digestion protocol used at PROBE can be found at (http://www.uib.no/filearchive/in-solution-proteindigestion.pdf

2.3.2 In gel protein digestion using trypsin followed by extraction of liberated peptides

Following protein separation using SDS-PAGE, the protein bands were excised from the gel, cut into 1mm cubes and transferred into individual Eppendorf tubes. 50-100 µL of wash solution (250 µL Ambic, 4750µL Milli-Q water and 5 mL Acetonitrile (ACN)(MS grade)) was added to each sample to cover the gel cubes, and incubated at room temperature for 20 minutes in Eppendorf mixer. The supernatant was removed and the washing step was repeated several times until the gel pieces were properly de-stained for Coomassie.

The gel pieces were further dried in a vacuum concentrator at 30 °C.
Trypsin (6 ng/ µL) Porcine was prepared by mixing 10 µL Trypsin Porcine (100 ng/µL dissolved in 50 mM acetic acid) with 160 µL digestion buffer (1M Ambic and 50 µL ACN to 900 µL deionized water), 20-40 µL was added to each sample and hydrated on ice for 30 minutes. The samples were further incubated for 16 h at 37 °C for protein digestion. The samples containing tryptic peptides were cooled in room temperature, spun down and the supernatant was transferred to a fresh tube. The remaining gel pieces were further extracted by addition of 30-50 µL 1% trifluoroacetic acid (TFA) (Sigma-Aldrich, St. Louis, MO, USA), and incubated in room temperature for 20 minutes with gentle shaking. The supernatant was pulled off and pooled with the first extraction. A final extraction with 30-50 µL 60% ACN/ 0.1% TFA was pooled with the former extractions. The sample with extracted peptides was concentrated to 10-15 µL in a vacuum concentrator.

The protocol for in-gel digestion is available at http://www.uib.no/filearchive/in-gel-proteindigestion.pdf)

2.4 Stable isotope dimethyl labeling and spike in of stable isotope labeled internal standards

2.4.1 Stable isotope dimethyl labeling

Stable isotope dimethyl labeling was performed as described in (Hsu, Huang et al. 2003; Boersema, Aye et al. 2008). In brief, desalted CSF sample containing tryptic peptides were dissolved in 100 µL of 100 mM triethylammonium bicarbonate (TEAB) (Sigma-Aldrich, St.Louis, MO, USA). By addition of 10 µL 4% CH_2O, 4% CD_2O and 4% $13CD_2O$ (all from Isotec™, Miamisburg, OH, USA), a reference pool was labeled light, whereas the individual control and case samples were labeled intermediate and heavy dimethyl label, respectively. The reducing agents were added as follows; 10 µL of 600 mM sodium cyanoborohydride ($NaBH_3CN$) (Fluka Analytical, Buchs, Switzerland) was added to the light and intermediate labeled samples while 10 µL of 6 mM cyanoborodeuteride ($NaBD_3CN$) (Isotec™, Miamisburg, OH) was added to the heavy labeled samples followed by 2 h incubation at room temperature with gentle shaking (350 rpm). The reaction was quenched with 16 µL 1% ammonium solution. Finally, 8 µL 100% FA was added on ice as a subsequent quenching step and all three labeled samples (light, intermediate and heavy labeled) were pooled. The pooled samples was subjected to a second step of desalting and subsequently stored at -20 for further purification and mass spectrometric analysis.

2.4.2 Stable isotope labeled internal standards (SISs)

Stable Isotope Dilution Selected Reaction Monitoring (SID-SRM) allows for relative and accurate quantification of peptides and proteins, and involves spiking the samples with synthetic peptides incorporated with stable isotope, termed stable isotope labeled internal standards (SISs).

Selection and optimization of stable isotope labeled internal standards

Signature peptides for the biomarker candidate proteins were defined and their uniqueness examined with a BLAST search against all human proteins in the NCBI database. SISs incorporated with ^{13}C and ^{15}N corresponding to the signature peptides were purchased in crude quality from Thermo Scientific. Ideally, the SISs peptides should be added to the sample in amounts corresponding to the endogenous peptide levels, at least within one order of magnitude in concentration. For this purpose SISs were optimized by direct infusion on a Q-Trap 5500 (AB SCIEX, Framingham, MA, USA) and the three most intense fragments were included in the final SRM-assay. A standard curve was drawn to find the concentration where of the SIS peptide are in a 1:1 relative level to the target endogenous peptide. Signature peptides were defined and SISs were optimized by others at PROBE.

SISs were spiked into trypsin digested CSF samples prior desalting with C18 StageTip (3M Empore™) as described in section 2.5.1.

2.5 Sample purification prior Mass spectrometry

Mass spectrometers are highly sensitive instruments; to avoid undesired signal interference and clogging of the pre-column sample clean up is performed prior injection into the instruments.

2.5.1 C18 StageTip

Two layers of C18 StageTip (3M Empore™, Eagan, MN, USA) were packed into each pipette tip, serving as a column. The column was conditioned with 20 µL 60% ACN/ 0.1% FA, followed by 40 µL 0,1% FA. The concentrated CSF samples containing tryptic peptides were dissolved in 20 µL 0.1% FA and added to the column, followed running through the column with 40 µL 0.1% FA. The peptides were eluted by addition of 10 µL 60% ACN/ 0.1% FA followed by a last addition of 1 µL 100% ACN to certify that every peptide was eluted. In order to exclude the high concentration of ACN the samples were dried in a vacuum concentrator. Followed by resolving in 0.5 µL 100% FA and 19.5 µL 3% ACN/ 5% FA.

2.5.2 C8 StageTip

The C8 filters (3M Empore™, Eagan, MN, USA) were conditioned with 10 µL 80% ACN/0.1% FA. Concentrated peptide samples were dissolved in 0.5 µL 100% FA and 9.5 µL 0.1% FA and run through the filter. Followed by a second addition of 5 µL 80% ACN/0.1% FA to ensure that all peptides were eluted form the column. Subsequently the samples were concentrated to dryness in a vacuum concentrator followed by a resuspension in 0.1% FA.

2.5.3 Oasis® HLB µElution Cleanup

Digested and concentrated CSF samples containing tryptic peptides were resuspended in 0.1% FA and desalted using reverse phase Oasis® HLB µElution Plate 30 µm (Waters, Milford, MA, USA). In brief, the plate were washed once with 500uL 80% ACN/0.1% FA, and subsequently washed twice with 500uL 0,1% FA. The dissolved samples were added to the wells on the µElution plate followed by washing three times with 500uL 0,1% FA. At this point the peptides are bound to the filters; to elute the bound peptides 100uL of 80% ACN/0.1% FA was added twice. The peptide samples were then transferred to fresh tubes and dried in a vacuum centrifuge and stored at -80^{o}C until further analysis. The centrifuge speed was 200 *x g* for one minute for all steps, except the steps concerning sample addition, where 150 *x g* was used for three minutes. Centrifuge used for Oasis® HLB µElution cleanup was Biofuge Statos (Heraeus, Buckinghamshire, USA).

2.6 Development of SRM blood contamination assay

SISs corresponding to signature peptides for hemoglobin were purchased from Thermo Scientific in crude quality. Each peptide was manually optimized by direct infusion on a Q-Trap 4000 (AB SCIEX, Framingham, MA, USA). The three most intense transitions were selected and optimized with respect to collision energy (CE), declustering potential (DP) and collision cell exit potential (CXP).

CSF test sample (i.e. pooled sample with CSF from multiple patients with various neurological conditions) were spiked with 1 µL blood/mL CSF. Each sample was

reduced, alkylated and trypsin digested (Section 2.3.1). SISs representing hemoglobin beta-1 was spiked into the CSF test samples at different concentration ranging from 100 amol to 1 pmol, prior to desalting using C18 StageTip (Section 2.5.1). Each sample was subjected to LC SRM-MS analysis on a Dionex Ultimate NCS-3000 nano-LC system coupled to a Q-Trap 4000. The amount of CSF loaded onto the column was kept constant (1µg CSF protein digest for each SISs dilution). The area ratio was calculated by dividing the area of SIS on the area of endogenous peptide, as the concentration of endogenous peptides was kept constant. The area ratio was plotted against the different SIS concentrations for generation of response curves. The approximate 1:1 ratios of the SISs and endogenous peptide were determined to generate a hemoglobin SRM-assay in order to detect hemoglobin in the CSF samples.

The SRM data was analyzed using MultiQuant 2.1.1 (AB SCIEX, Framingham, MA, USA). SISs and endogenous CSF hemoglobin peptides were manually integrated and the data was evaluated in which transitions from the endogenous peptides with wrong retention time, poor integrations and signal-to-noise less than 3 were considered to be free of signal from hemoglobin e.g. free of blood contamination.

2.7 Mass spectrometric analysis

2.7.1 Nano-HPLC ESI-LTQ Orbitrap Velos Pro mass spectrometry

For LC-MS/MS analysis desalted samples containing tryptic peptides was resuspended in 0.1 % FA and 0.5 µg/µL was injected onto a pre-column (Dionex, Acclaim PepMap Nano Trap column, C18, 75 µm i.d. x 2 cm, 3 µm) followed by separation on the analytical column (Dionex, Acclain PepMap100 RSLCnano column, 75 µm i.d. x 15cm, C18, 2 µm) using a Dionex Ultimate NCS-3500RS LC system coupled online to an Orbitrap Velos Pro (Thermo Scientific, Waltham, MA, USA) mass spectrometer. The LC run was of 90 minute (180 minute) duration and peptides were separated during a biphasic ACN gradient from two nanoflow UPLC pumps (flow rate of 280 nL /min) on the analytical column. Solvent A and B was 0.1% FA (vol/vol) with 2% ACN and 90% ACN (vol/vol) respectively. The gradient composition was 8-38% B over 61.5 min, then 38-90% B over 3 minutes. Elution of

very hydrophobic peptides and conditioning of the column were performed during 5 minutes isocratic elution with 90% B and 12 minutes isocratic elution with 5% B respectively.

The eluting peptides were ionized in the electrospray and analyzed by the Orbitrap Velos Pro. The mass spectrometer was operated in the DDA-mode (data-dependent-acquisition) to automatically switch between full MS scan in the Orbitrap and MS/MS acquisition in the ion trap. The instrument control was performed through Tune 2.6.0 and Xcalibur 2.1. Survey full scan MS spectra (from m/z 300 to 2000) were acquired in the Orbitrap with resolution R=60,000 at m/z 400 (after accumulation to a target value of 1e6 in the linear ion trap with maximum allowed ion accumulation time 500ms). The Top 7 (or 15) most intense eluting peptides above a ion threshold value of 1000 counts, and charge state 2 or higher, were sequentially isolated to a target value of 1e4 and fragmented in the high-pressure linear ion trap by low-energy CID (collision-induced-dissociation) with normalized collision energy of 35% and wideband-activation enabled. The maximum allowed accumulation time for CID was 200 ms, the isolation width maintained at 2Da, activation q=0.25, and activation time of 10ms. The resulting fragment ions were scanned out in the low-pressure ion trap at normal scan rate, and recorded with the secondary electron multipliers. One MS/MS spectrum of precursor mass was allowed before dynamic exclusion for 30 sec.

2.7.2 Inclusion list dependent acquisition on the Orbitrap Velos Pro MS

The following parameters were used for inclusion list dependent acquisition on the Orbitrap Velos Pro mass spectrometer. A single Orbitrap MS scan from m/z 300 to 1500 at resolution 60,000 was followed by four ion trap MS/MS scans at normal scan rate. The top four most abundant precursors from the inclusion list were targeted for MS/MS spectrum acquisition over the course of a 90-min experiment. Preview mode and charge state screening were enabled for selection of precursors. The m/z tolerance around targeted precursors was +/- 7.5 ppm. Dynamic exclusion was also enabled with a repeat count of 1.5 with a repeat duration of 6 s and exclusion duration of 30 s. Again the m/z tolerance for dynamic exclusion was 7.5 ppm. The intensity threshold

for triggering of a detected peak was set to 250, and collision energy was specified at 40% for all inclusion list members.

2.7.3 Stable Isotope Dilution SRM-MS analysis

For LC SRM-MS analysis a Q-Trap 5500 coupled to a Dionex Ultimate 3000RS nano-LC system was used. The concentrated samples were resuspended in 3% ACN/ 5% FA, and 1 µL, corresponding to 1 µg CSF protein digest, was loaded onto the pre-column (Dionex, Acclaim PepMap Nano Trap column, C18, 75 µm i.d. x 2 cm, 3 µm) followed by separation on a custom made column (Dionex, Acclain PepMap100 RSLCnano column, 75 µm i.d. x 15 cm, C18, 2 µm). The LC analysis was of 70 minute total duration with the following mobile phase: mobile phase A (0.1% FA) and mobile phase B (0.1% FA/ 90% ACN). The gradient used was as follows: 5-10 % B from 0-3 minutes, 10-45 % B from 3-45 minutes and 45-90% B from 45-45.5 minutes, and hold at 90% B from 45.5-51.5 minutes. From 51.5-55 minutes ramp from 90-5% B and re-equilibrate column at 5% B from 55-70 minutes. The flow rate was 250 nL/min. The MS run time was 68 minutes. The peptides were analyzed using scheduled SRM, the target scan time was 1 second and the detection window 240 seconds. Three transitions per peptide were monitored, whereof one transition per peptide was selected for accurate relative quantification analysis. Supplementary table 6.6 and 6.7 displays the signature peptides, transitions used quantification, Q1 and Q3 m/z values for the endogenous and SIS peptides and the collision energy used in the SRM analysis.

2.8 Bioinformatics

2.8.1 Protein identification using SearchGUI/ Peptide Shaker or Spectrum Mill

The raw data files from MS analysis were search with both Sepctrum Mill (Agilent Technologies, Santa Clara, CA, USA) and SearchGUI and Peptide Shaker. (SearchGUI combine the search engines OMSSA and X!Tandem). The settings for protein identification were as follows; fixed and variable modifications were carbamidomethyaltion (c) and oxidation of methionine (m). Protease was trypsin,

Precursor Ion Mass Tolerance 10 ppm, Fragment Ion Mass tolerance 0.7 Da, fragment ion type 1: b and fragment ion type 2: y. Minimum precursor charge was 2, max missed cleavages was 2. Max precursor charge 4. The peptides and proteins identified by the search was autovalidated using the default settings and the false discovery rate (FDR) of less than 1%. The human Swissprot database (European Bioinformatics Institute (EBI) and the Swiss Institute of Bioinformatics (SIB)) was used.

2.8.2 Progenesis LC-MS label-free proteomics analysis

The software Progenesis LC-MS (Nonlinear Dynamics, Newcastle upon Tyne, UK) was used for label-free quantification and comparison of LC-MS proteomics data based on the volume of MS^1 peaks. The MS^1 features from the individual LC-MS runs were represents peptides and were presented in a 2D plot with m/z versus retention time. A feature is a peptide with its known retention time, m/z, isotopic distribution and volume (intensity). The intensity information was used to determine the relative quantitation of the peptides. The different runs were aligned based on retention time in order to compensate for any retention time drift between the runs. This enabled the detection of peptides with their differential abundance across the compared groups. Peptides were filtered and only peptides with a minimum fold change 1.5 between the compared groups and an analysis of variance (anova) p-value of less than 0.05 were selected for identification. The associated MS/MS spectra of the selected peptides were search against the human SwissProt database using SearchGUI, in combination with Peptide Shaker for protein identification as described above.

2.8.3 Generation of inclusion list for AIMS

In the AIMS approach, 17 proteins ranging from high to median abundance in CSF were targeted. For high abundant proteins five peptides per protein were selected while 10 peptides were targeted for lower abundant proteins. To generate the inclusion list, which is a key factor in targeted AIMS approach, the accession number of the proteins was searched in Expasy.org Bioinformatics Resource Portal using the database UniProtKB. From UniProtKB, the FASTA file for each protein was exported

into MSDigest (University of California, San Francisco), were in silico digestion was performed. The selected peptide criteria were as follows: proteins with max missed cleavages 0, constant modification carbamidomethyl (C), peptide mass (m/z) of 600.0 to 4000.0. The MS-Digest generates the peptide sequence with monoisotopic mass of singly charge. These masses were further used to calculate the +2 and +3 mass values. The m/z of the peptide at a given charge is between 300 and 1500.

The inclusion list for stable isotope dimethyl labeled peptides were generated by adding up the 28 Da (light labeled version), 32 Da (intermediate labeled version) and 36 Da (heavy labeled version) to the original non-labeled mass of respective peptides as shown in Table 2.1.8. The final inclusion list of 354 selected peptides are available in Supplementary Table 3.2 B.

Lysine (K)				Arginine (R)			
+2 L	+28.0313	+3 L	18.6875	+ 2 L	+14.0156	+ 3 L	+9.3437
+2 IM	+32.0564	+3 IM	21.3709	+ 2 IM	+16.0282	+3 IM	+10.6854
+2 H	+36.0756	+3 H	24.0504	+ 2 H	+18.0378	+3 H	+12.0252

Table 2.1.8: Generation of inclusion list for AIMS on stable isotope dimethyl labeled peptides. Dependent on if the peptide sequence ended with lysine (K) or arginine (R) the 2+ and 3+ mass values was calculated for light (L), intermediate (M) and heavy (H).

2.8.4 Analysis of dimethyl-AIMS data using Proteome Discoverer Daemon

The raw data from the inclusion list dependent acquisition on the Orbitrap Velos Pro mass spectrometer was analyzed using Proteome Discoverer Daemon (Thermo Scientific, Waltham, MA, USA). The MS/MS data was used for protein identification, while the MS data was used for quantification.

MS/MS spectra were extracted from the individual LC-MS/MS analysis and subsequently searched against the human SwissProt database. The MS/MS data were searched in Mascot in Proteome Discoverer in which the following parameters were used: 50 p.p.m., precursor mass tolerance, 0.5 Da fragment ion tolerance, no-missed cleavages, carbamidomethyl cysteine as fixed modification and oxidized methionine

as variable modifications. A dimethyl-based quantitation method was chosen in Proteome Discoverer, with mass precision requirement of 2 p.p.m. for consecutive precursor measurements. Taking into account the isotopic effect of deuterium, 1 min of retention time tolerance was applied for isotope pattern multiplets and allowed spectra with 2 missing channels to be quantified. After identification and quantification, the data was filtered according to very strict peptide acceptance criteria. These criteria included high confidence level and position rank 1 in Mascot search.

2.8.5 Analysis of SRM data by use of Analyst and MultiQuant

The mass spectrometric method SRM involves two stages of mass filtering. In the first mass filter, the target peptide (precursor ion) representing the target protein is selected and isolated followed by fragmentation. In the second mass filter, predefined fragments of the peptide ion (i.e. fragment ion) are monitored. The combination of precursor and the respective fragment ion is referred to as a transition and selected transitions for each peptide are used to monitor the target proteins across samples. For accurate relative quantification synthetic peptides, stable isotope–labeled internal standards (SISs) are introduced at known concentrations to each sample. The stable isotope labeled peptide will have identical properties as its endogenous counterpart, except the mass will be higher due to the synthetic peptide being enriched with stable isotopes. This mass increase allows the two peptides to be distinguished during analysis. Abundance of the target protein is determined by comparing the signal from the SIS peptide and the endogenous peptide in the sample.

The raw data files from the SRM LC-MS were analyzed using the software Analyst® 1.5.1 and MultiQuant™ 2.1.1 (both from AB SCIEX, Framingham, MA, USA). Analyst® is used to display the chromatograms, while MultiQuant™ enable peak integration. The selection of peaks is based on retention time. Transitions of where the endogenous peptide differed by more than 0.05 in retention time from the SIS peptide were excluded together with transitions displaying twin peaks. By performing peak integration the area ratio was obtained, which is the area under the curve of the endogenous peptide over the area under the SIS peptide. The area ratio is used to

quantify the differential abundance of target proteins between samples. In this manner, the area ratio was used to measure the differential abundance between the endogenous and SIS peptide. The average area ratio of transitions was calculated and used for quantification between samples.

3 Results

Proteomics-based biomarker discovery in multiple sclerosis

The schematic flow chart summarizes the main steps performed in the proteomics-based biomarker discovery and verification studies presented in this section. In the screening phase we combined a CSF pooling strategy and a label-free relative quantitation approach to reveal proteins of differential abundance in MScl compared to controls. A selection of these proteins was further evaluated as biomarker candidates in MScl, by performing two independent quantification strategies. In the verification phase we performed biomarker qualification using a dimethyl-AIMS approach as an initial verification strategy to evaluate the potential of the proteins found in the screening phase as biomarkers candidates in MScl. A selection of potential biomarker candidates from the discovery study was supplemented with candidate proteins derived from the literature and subjected to biomarker verification using SID-SRM analysis.

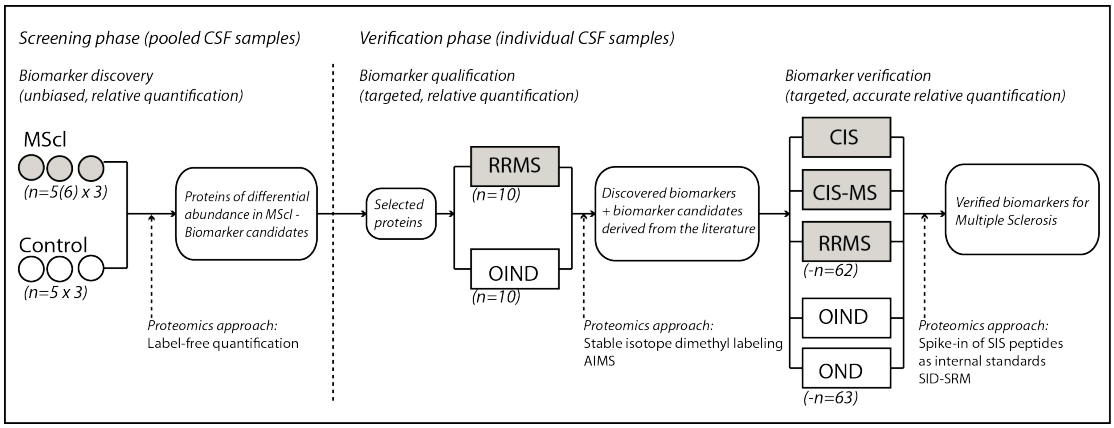

Figure 3.1.1: **Biomarker discovery, qualification and verification in multiple sclerosis.** The flow chart summarizes the different steps undertaken in the biomarker pipeline. In the first phase of the study (*screening phase*), a mass spectrometry-based proteomic approach (label-free quantification) was applied to pooled CSF samples form combined* MScl patients (n=16) and combined controls (n=15), in order to identify and quantify proteins of differential abundance (biomarker candidates) in CSF of MScl patients. In the following phase (*verification phase*), a selection of the biomarker candidate proteins showing significant up-or down regulation in MScl, were chosen for qualification in 10 novel individual RRMS patients and 10 novel OIND as controls. Stable isotope dimethyl labeling of the samples was combined with the mass spectrometry-based method Accurate Inclusion Mass Screening (AIMS). Potential biomarker candidates were supplemented with biomarker candidates derived from the literature and targeted for verification in a larger patient cohort, by incorporation of broader range of patients (n=62) and controls (n=63). Biomarker verification was performed using Stable Isotope Dilution Selected Reaction Monitoring (SID-SRM) for accurate relative quantification. *Combined MScl (13 RRMS, 1 CIS-MS,1 SPMS and 1 PPMS) and combined controls (10 OIND and 5 OND).

3.1 Discovery of differential abundant proteins in multiple sclerosis using label-free relative quantification

The aim of the biomarker discovery was to compare the CSF proteome of combined multiple sclerosis (MScl) patients with the CSF proteome of other related disease groups as control, in order to identify and quantify proteins of differential abundance between the groups.

3.1.1 Included patients and analytical approach

All CSF samples (n=31) included in this study were considered free for blood contamination based on the criteria described in Section 2.6 in Materials and Methods. Each CSF sample was checked for blood contamination in a similar manner as described in the same section.

By combining a CSF pooling strategy and a label-free quantification approach, relative quantitation of differentially abundant proteins between MScl (n=16) and controls (n=15) were obtained. All patient information are available in Supplementary Table 6.1. A total of six CSF pools were created, three MScl pools and three control pools. Each pool contained CSF from five different patients, whereof each patient contributed with 20 µg to a final amount of 100 µg per pool. One of the MScl pools consisted of six patients. The control groups consisted of patients with other inflammatory neurological diseases (OIND) and other neurological diseases (OND).

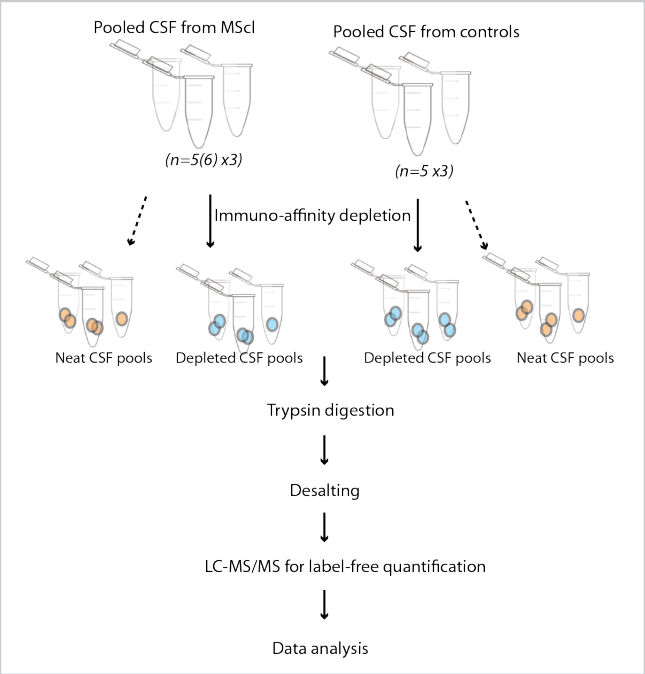

Figure 3.1.2: Biomarker discovery study design. The three MScl pools and three control pools were each processed on two different levels prior LC-MS/MS on an Orbitrap Velos Pro MS for label-free quantification. 95 % of the total protein amount of each CSF pool was depleted for 14 high abundant proteins, while the remaining 5% was trypsin digested directly. The portioning of each pool enabled biomarker discovery in both neat and depleted CSF. Data analysis was performed using Progenesis LC-MS.

Each of the six CSF pools was processed in two different ways (as illustrated in Figure 3.2.1). In one approach, 95 % of the total sample was subjected to immuno-affinity depletion prior trypsin digestion (Section 2.3.1 in Materials and Methods), while in the other approach the remaining 5 % of each sample was trypsin digested directly, without depletion (Section 2.2.2).

Each of the 12 generated CSF samples, 6 neat CSF pools and 6 depleted CSF pools, each comprising approximately 5 µg protein, was trypsin digested (Section 2.3.1) and desalted (Section 2.5.3). Dried samples were resuspended in 0.1% FA and 0.5µg CSF protein digest from each of the 12 samples were subjected to LC-MS/MS using a Dionex Ultimate 3000RS nano-LC system coupled to an Orbitrap Velos Pro MS. The mass spectrometer automatically switched between full MS scans in the Orbitrap and MS/MS acquisition in the ion trap in a DDA-mode (data-dependent-acquisition). The LC analysis was of 90 minute duration and the top 7 most intense eluting peptides were sequentially isolated and fragmented in the high-pressure linear ion trap by CID (collision-induced-dissociation) as described in Section 2.7.1.

The data resulting from the MS analysis was further analyzed using the Progenesis LC-MS software (Section 2.8.2). After using this software the abundance of proteins identified from the three MScl pools was compared to the corresponding protein abundance from the control pools. The data obtained from neat CSF samples of MScl patients were compared with the corresponding data obtained from controls. The data from the depleted CSF proteomes were also compared between the two groups. In this manner the data obtained from the neat and depleted CSF samples was analyzed separately in two different experiments.

3.1.2 Protein identification and statistical analysis for relative quantification

Analyzing the protein identification in neat CSF from both patient groups gave 250 protein identifications, while combining protein identification from MScl and controls in depleted CSF lead to identification of 410 proteins, hence 60% more proteins were identified after depleting the sample for the 14 high abundant CSF proteins. Supplementary Table 3.1 B displays proteins identified in neat and depleted CSF and their regulation level.

Proteins were considered as increased in relative abundance if the fold change was >1.5 between MScl and control, with an anova p-value of 0.05 or lower. Based on the set criteria, a total of 65 proteins with different relative abundance were discovered (Supplemenary Tabel 3.1 B). Of the 65 differentially abundant proteins, 29 were significantly regulated only in neat CSF, while 28 proteins were significantly regulated only in depleted CSF (Supplementary Table 3.1 B). As shown in the Venn diagram in Figure 3.1.3, four proteins were regulated in both CSF fractions, and these proteins were: Ceruloplasmin, Complement C1q subcomponent subunit C, Complement factor B and Ectonucleotide pyrophosphatase. All four proteins showed the same level of regulation in both neat and depleted CSF (Supplementary Table 3.1 B).

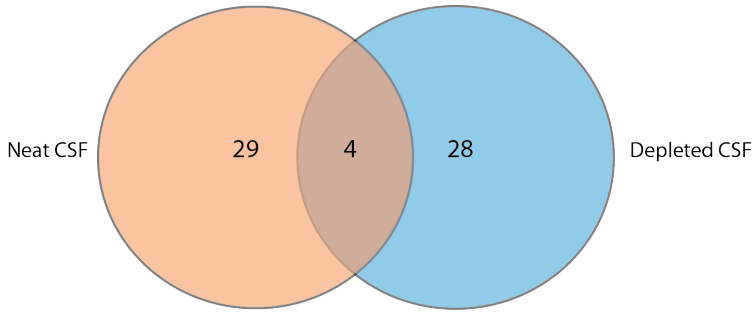

Figure 3.1.3: Proteins regulated in neat and depleted CSF. Of the 65 differentially abundant proteins, 29 were significantly regulated only in neat CSF, while 28 proteins were significantly regulated only in depleted CSF**.** Four proteins were found to be regulated in both CSF fractions: Ceruloplasmin, Complement C1q subcomponent subunit C, Complement factor B and Ectonucleotide pyrophosphatase. The proteins showed similar levels of regulation in neat and depleted CSF.

3.1.3 Biomarker candidate selection

Among the differentially abundant proteins discovered in neat CSF, six proteins were selected for further evaluation as biomarker candidates in MScl. While nine proteins derived from depleted CSF were selected. Ceruloplasmin and Ectonucleotide pyrophosphatase were discovered in both neat and depleted CSF. Hence, of the 65 proteins, 17 potential biomarker candidates were selected for further qualification in individual CSF samples. Selection of candidate biomarkers was based on previous implications in the literature of their role in MScl or other neuro-inflammatory and neuro-degenerative processes and proteins showing high fold changes between MScl and controls. Table 3.1.1 summarizes the 17 selected biomarker candidates; the peptides used for quantification their abundance level in MScl.

Protein	Acc. number	Peptides used for quantification	Peptide count	Spectral count	Confidence score	Anova p-value	Fold change	Level in MScl
Complement C3	P01024	LSINTHPSQKPLSITVR IHWESASLLR ISLPESLKR FISLGEACKK SEETKENEGFTVTAEGK LDKACEPGVDYVYK SGQSEDRQPVPGQQMTLK AKDQLTCNKFDLK AAVYHHFISDGVRK	9	65	992	0.022	-2.2	↓
Transthyretin	P02766	VLDAVRGSPAINVAVHVFR GPTGTGESKCPLMVK VLDAVRGSPAINVAVHVFRK	3	6	359	0.004	2.3	↑
Haptoglobin	P00738	LRTEGDGVYTLNNEK LRTEGDGVYTLNDKK	2	8	200	0.028	-3.8	↓
Alpha-2-macroglobulin	P01023	TEVSSNHVLIYLDK YSDASDCHGEDSQAFCEK	2	7	200	0.034	-1.7	↓
Alpha-1-acid glycoprotein 1	P02763	SDVVYTDWKK	1	6	100	0.039	-3.0	↓
Alpha-1-antitrypsin	P01009	GKWERPFEVK LGMFNIQHCKK	2	10	193	0.006	-3.3	↓
Semaphorin-7A	O75326	DCENYITLLER LQDVFLLPDPSGQWR YYLSCPMESR	3	19	300	0.024	-1.5	↓
Osteopontin	P10451	KANDESNEHSDVIDSQELSK NDESNEHSDVIDSQELSK	2	12	300	0.025	1.6	↑
Amyloid beta A4 protein	P05067	MDVCETHLHWHTVAK	1	4	200	0.040	1.6	↑
Prostaglandin-H2 D-isomerase	P41222	MATLYSR	1	6	100	0.002	-1.5	↓
Neuroserpin	Q99574	MAVLYPQVIVDHPFFFLIR	1	10	100	0.047	1.7	↑
Cystatin-C	P01034	QIVAGVNYFLDVELGR	1	1	100	0.026	3.7	↑
Pigment epithelium-derived factor	P36955	IAQLPLTGSMSIIPFLPLK EIPPEISILLLGVAH	2	7	200	0.004	1.7	↑
Zinc-alpha-2-glycoprotein	P25311	KSQPMGLWR AKAYLEEECPATLR	2	9	200	0.034	-1.9	↓
Hemopexin	P02790	GDKVWVYPPEKK DVRDYFMPCPGR QGHNSVFLIKGCK EWFWDLATGTMKER QGHNSVFLIKGCK	5	71	799	0.022	-1.9	↓
Ceruloplasmin	P00450	NLASRPYTFHSHGITYYK	1	5	100	0.043	-1.8	↓
Ectonucleotid Pyrophosphatase	Q13822	VRDIEHLTSLDFFRK	1	6	180	0.033	2.0	↑

Table 3.1.1: Selection of biomarker candidates identified and quantified with differential abundance levels in CSF pools of MScl patients versus controls. The table summarizes and lists the 17 biomarker candidates selected for further evaluation in the biomarker pipeline. CSF pools of MScl and controls were processed in two levels in order to enhance protein coverage of the CSF proteome. 5 % of each pool was directly trypsin digested, whereas the remaining 95 % were subjected to immuno-affinity depletion of 14 high abundant proteins in CSF. The first six proteins in the table represent biomarker candidates detected in neat CSF, the next nine proteins were discovered in depleted CSF, whereas the two latter proteins represent biomarker candidates discovered in both neat and depleted CSF pools of MScl and control patients. All selected biomarker candidates showed significant differential CSF levels with the fold change exceeding 1.5 between MScl and controls and an anova p-value less than 0.05.

3.1.4 Biomarker candidates discovered in neat CSF

Of the six biomarker candidates selected from the neat CSF, Haptoglobin, Alpha-1-acid glycoprotein 1, Alpha-1-antitrypsin, Ceruloplasmin, Complement C3 and Alpha-2-macroglobulin showed decreased CSF levels in MScl patients compared to controls, while Transthyretin and Ectonucleotid pyrophosphatase showed increased CSF levels in MScl. Figure 3.1.4 and 3.1.5 shows the abundance profiles of the selected proteins discovered in neat CSF.

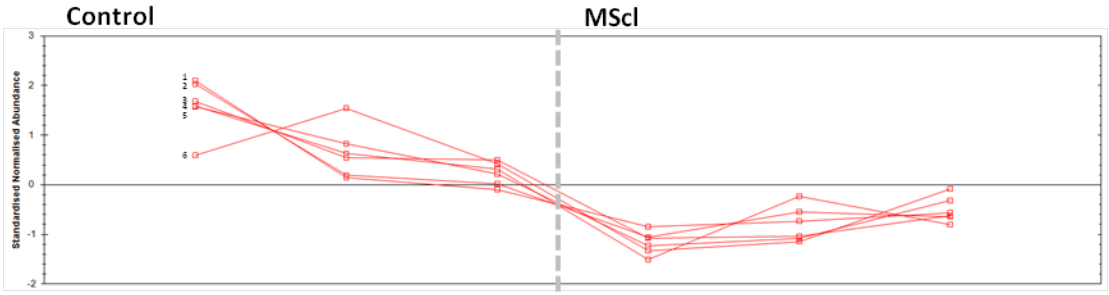

Figure 3.1.4: Normalized abundance profiles of proteins showing significant decreased abundance in neat CSF of MScl patients. By using the software Progenesis LC-MS standardized normalized abundance profiles of the six CSF pools were obtained. In MScl there were significant decrease in CSF levels of six proteins: 1. Haptoglobin, 2. Alpha-1-acid glycoprotein 1, 3. Alpha-1-antitrypsin, 4. Ceruloplasmin, 5. Complement C3 and 6. Alpha-2-macroglobulin.

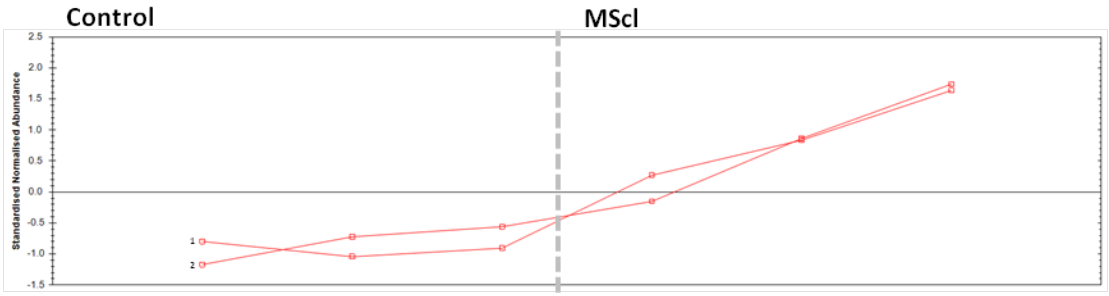

Figure 3.1.5: Normalized abundance profiles of protein showing significant increased abundance in neat CSF of MScl patients. The CSF levels of Transthyretin and Ectonucleotid pyrophosphatase were significantly increased compared to CSF levels in controls. 1. Transthyretin and 2. Ectonucleotide pyrophosphatase.

3.1.5 Biomarker candidates discovered in depleted CSF

Depleting the samples of 14 abundant CSF proteins enabled the discovery of 32 proteins with significant differential CSF levels in MScl patients compared to controls. Of these nine proteins were selected for further evaluation as biomarker candidates in MScl. Hemopexin, Semaphorin-7A, Zinc-alpha-2-glycoprotein and Prostaglandin-H2 D-isomerase were found with significant decreased CSF levels in MScl patients relative to the CSF levels in controls. Figure 3.1.6 and 3.1.7 shows the abundance profiles of the selected proteins discovered in depleted CSF.

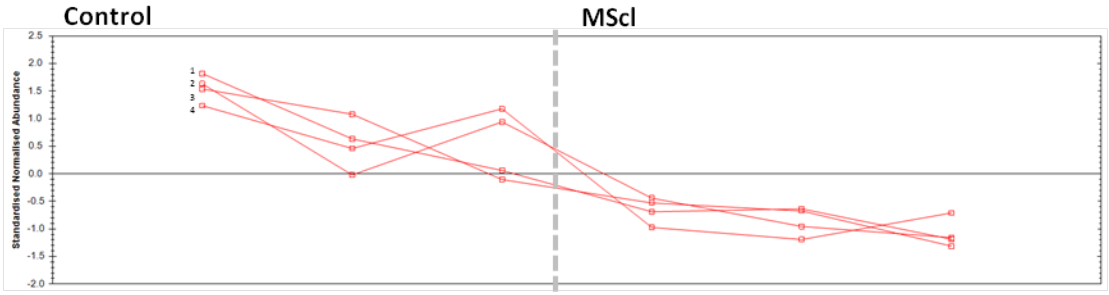

Figure 3.1.6: Normalized abundance profile of proteins showing significant decreased abundance in depleted CSF of MScl patients. 1. Hemopexin, 2. Semaphorin-7A, 3. Zinc-alpha-2-glycoprotein, 4. Prostaglandin-H2 D-isomerase.

Amyloid beta A4, Osteopontin, Cystatin-C, Pigment epithelium-derived factor and Neuroserpin were significantly increased in CSF levels of MScl patients. The abundance profiles of these five proteins are shown in figure 3.1.7.

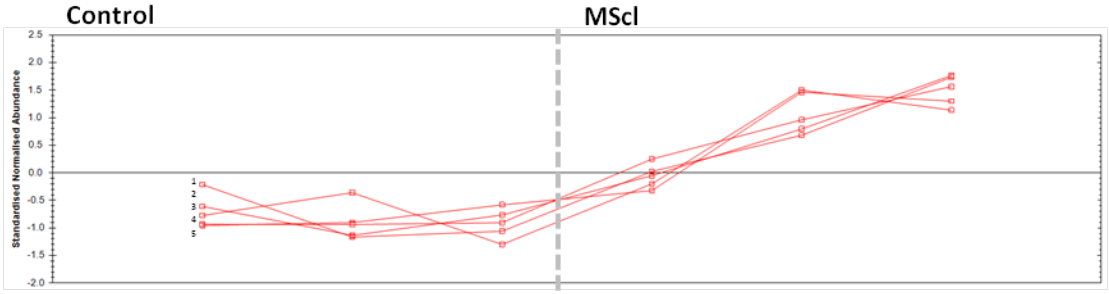

Figure 3.1.7: Normalized abundance profiles of protein with significant increased abundance in depleted CSF of MScl patients. 1. Amyloid beta A4, 2. Osteopontin, 3. Cystatin-C, 4. Pigment epithelium-derived factor and 5. Neuroserpin.

Verification of differential abundance of biomarker candidates

The false discovery rate was expected to be rather high for the biomarker candidates identified in pooled CSF samples during the screening phase. Hence, a selection of the biomarker candidates derived from the discovery study was subjected to comprehensive evaluation in two independent verification studies; involving biomarker qualification and verification.

3.2 Biomarker qualification using Accurate Inclusion Mass Screening

Qualification of biomarker candidates represents an initial verification process to assess the potential of the biomarker candidates in novel patients using an alternative quantification strategy. The 17 biomarker candidates derived from the discovery experiment (Table 3.1.1) were targeted for relative quantification to see if their differential abundance remained detectable between RRMS and controls when analyzing individual CSF samples. We applied stable isotope dimethyl labeling coupled to Accurate Inclusion Mass Screening (AIMS) for targeted quantification of 17 proteins comparing individual CSF levels from 10 RRMS patients and 10 OIND.

3.2.1 Experimental design for dimethyl-AIMS analysis

5 µg CSF each from 10 individual RRMS patients and 10 individual OIND patients were included in the qualification experiment. 5 µg of a global internal standard (GIS) was created by pooling the samples from RRMS patients (n=10) and OIND patients (n=10). All CSF protein samples were trypsin digested (Section 2.3.1), desalted (Section 2.5.3) and dried prior to labeling with dimethyl reagents. The GIS, OIND (controls) and RRMS (patients) were labeled with light-, intermediate- and heavy-dimethyl reagent, respectively (Section 2.4.1). One portion of GIS (light version) was mixed with one sample from OIND (intermediate version) and one sample from RRMS (heavy version) generating 10 mixtures of samples. These resulting 10 pooled samples were desalted (Section 2.5.3), dried and resuspended in 0.1% FA prior analysis. 0.5 µg from each sample was analyzed by inclusion list dependent acquisition (AIMS) on the Orbitrap Velos Pro MS (Section 2.7.2).

An inclusion list was generated comprising ten unique peptides for each low abundant biomarker candidate protein, and five unique peptides representing the high abundant proteins. The criteria for selected peptides were as follows; peptides were required to be unique to each of the proteins of interest, fully tryptic with no missed cleavages, and have no known or predicted post-translational modifications. In total 354 peptides were targeted for dimethyl-AIMS analysis of the 17 biomarker candidates. The +2 and +3 charge state for each of the peptides was used to generate the inclusion for the targeted quantification analysis as described in Section 2.8.3. For AIMS analysis, m/z

values on the inclusion list (Supplementary Table 3.2 B) were monitored in each scan on a high mass resolution and mass accuracy MS system, and MS/MS spectra were acquired only when a peptide from the list was detected with both the correct mass and charge state.

The dimethyl-AIMS data was analyzed using Proteome Discoverer Daemon (Thermo Scientific) as described in Section 2.8.4. The MS/MS spectra were extracted from the individual LC-MS/MS analysis and subsequently searched using Mascot against the human SwissProt database. The MS data was searched in Proteome Discoverer in which a dimethyl-based quantification method was chosen. Following identification and quantification, the data was filtered according to peptide acceptance criteria including high confidence level and position mark 1 in the Mascot search.

3.2.2 Qualification of biomarker candidates

The peptide abundance of the biomarker candidates was monitored across 10 experiments. In each experiment the peptide abundance of one RRMS patient (heavy dimethyl labeled) and one OIND patient (intermediate dimethyl labeled) was compared with the peptide abundance of the pooled GIS (light dimethyl labeled). A Student's t-test p-value less than 0.05 was considered significant.

As shown in Figure 3.2.1, the CSF levels of Alpha-1-acid glycoprotein 1, Alpha-2-macroglobulin, Ceruloplasmin, Prostaglandin-H2 D-isomerase and Transthyretin obtained from the initial screening phase were all showing similar levels of regulation in individual CSF samples from novel patients. The differential abundance of Transtyretin was significant (p-value 0.007). The abundance levels of Cystatin C and Complement C3 contradicted the data obtained from the discovery experiment, however the abundance level of these two proteins obtained from the qualification experiment was supported by findings in the literature for Complement C3 (Stoop, Dekker et al. 2008) and Cystatin C (Qin, Qin et al. 2009), respectively. Hence, the first seven proteins shown in Figure 3.2.1 were further included in the list of biomarker candidates for further verification.

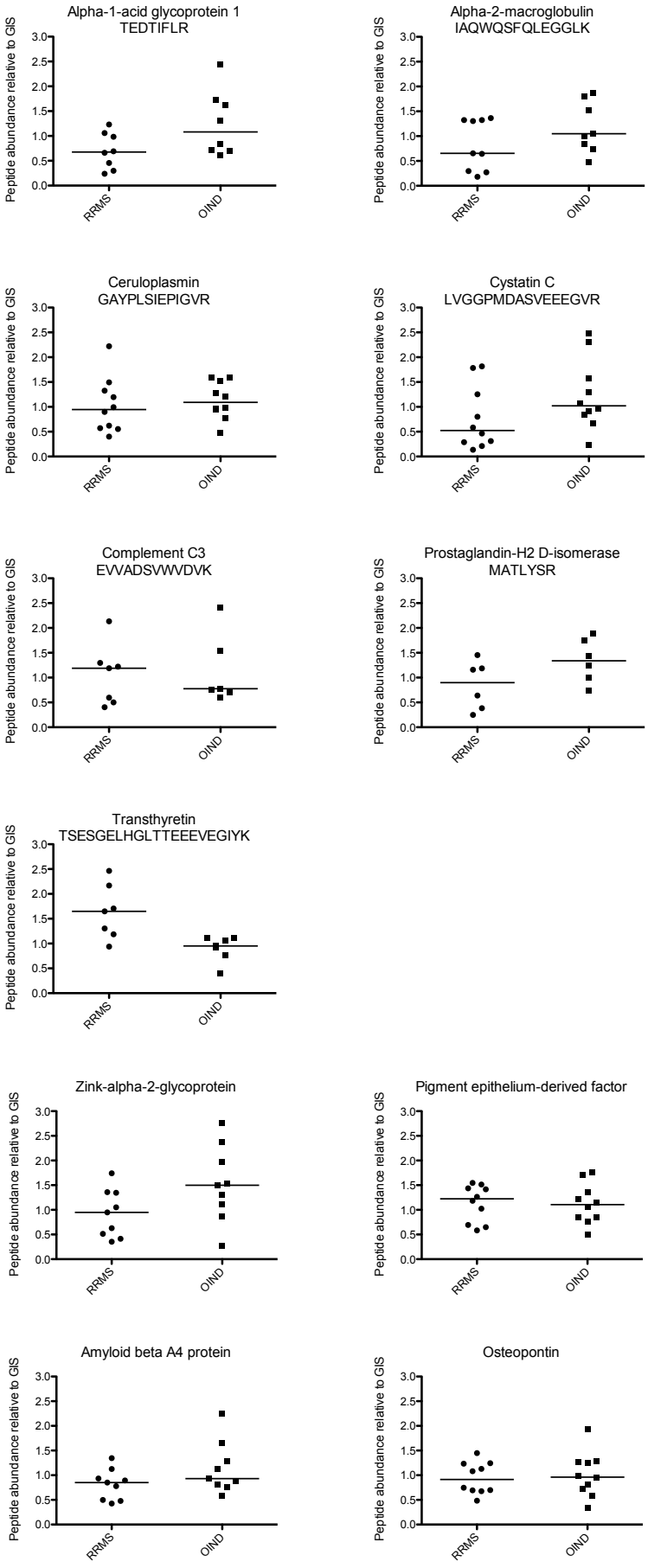

Figure 3.2.1: Abundance levels of biomarker candidates using dimethyl-AIMS. The figure displays the abundance level obtained from the qualification experiment for 11 proteins. Seven of these proteins were included in the SID-SRM verification study. By applying stable isotope dimethyl labeling coupled to the AIMS approach the peptide abundance of the biomarker candidate proteins were monitored across 10 experiments. The qualification experiment resulted in verification of the differential abundance level of five biomarker candidates derived from the discovery experiment. In each chart the median peptide abundance are shown for the experiments that an abundance ratio was obtained. The differential abundance ratio for Transthyretin was significant.

Protein	Peptide used for quantification	Median peptide abundance		Fold change (RRMS/OIND)	
		RRMS	OIND		
Alpha-1-acid glycoprotein 1	TEDTIFLR EQLGEFYEALDCLR SDVVYTDWK	0.68	1.08	-1.6	↓RRMS
Alpha-2-macroglobulin	IAQWQSFQLEGGLK AIGYLNTGYQR FEVQVTVPK QFSFPLSSEPFQGSYK	0.65	1.05	-1.6	↓RRMS
Ceruloplasmin	GAYPLSIEPIGVR ALYLQYTDETFR	0.95	1.09	-1.2	↓RRMS
Cystatin C	LVGGPMDASVEEEGVR ALDFAVGEYNK TQPNLDNCPFHDQPHLK	0.52	1.02	-2.0	↓RRMS
Complement C3	EVVADSVWVDVK	1.19	0.78	1.5	↑RRMS
Prostaglandin-H2 D-isomerase	MATLYSR AQGFTEDTIVFLPQTDK TMLLQPAGSLGSYSYR	0.90	1.34	-1.5	↓RRMS
Transthyretin	TSESGELHGLTTEEEFVEGIYK GSPAINVAVHVFR YTIAALLSPYSYSTTAVVTNPK	1.65	0.95	1.7	↓RRMS
Pigment epithelium-derived factor	IAAAVSNFGYDLYR TSLEDFYLDEER	1.22	1.10	1.1	↑RRMS
Zinc-alpha-2-glycoprotein	YSLTYIYTGLSK	0.95	1.50	-1.6	↓RRMS
Amyloid beta A4 protein	AVIQHFQEK VESLEQEAANER	0.85	0.93	-1.1	↓RRMS
Osteopontin	ISHELDSASSEVN	0.91	0.96	-1.1	↓RRMS

Table 3.2.1: Verification of differential abundance of biomarker candidate proteins. The table lists the seven proteins also included in the SID-SRM verification study, in addition an abundance ratio was obtained for the four latter proteins. All detectable peptides used for quantification are shown, in which the abundance level of the marked peptides are shown in Figure 3.2.1.

SRM-assays for seven of the proteins listed in Table 3.2.1 were already developed, and these proteins were therefore included in the SID-SRM verification study. The differential abundance of Pigment epithelium-derived factor, Osteopontin, Zinc-alpha-2-glycoprotein and Amyloid beta A4 protein obtained in the qualification study was not significant and as no SRM-assays were available for these proteins, they were discarded as biomarker candidates for further verification.

3.3 Verification of biomarker candidates using Selected Reaction Monitoring

Existing literature on proposed CSF biomarker candidates in MScl is extensive. However, many of the reported biomarker studies are performed in relatively small patient cohorts and lack further verification. Hence, in addition to eight biomarker candidates from the discovery experiment a selection of candidate proteins derived from the literature was included in a verification experiment using SID-SRM for targeted quantification. The literature-derived proteins were selected based on their reported differential abundance in MScl from various proteomics-based biomarker

discovery studies as well as indications in the literature of their role in MScl or other CNS inflammatory and degenerative processes.

3.3.1 Study design and conduction

The presented verification study included CSF samples from 125 patients in total. All patient information is available in Supplementary Table 6.3. CSF samples from 62 CIS or multiple sclerosis patients and 63 controls were included. 13 CIS, 14 CIS-MS and 35 RRMS patients were compared with 32 OIND and 31 OND. Including CSF from patients diagnosed with CIS at the time of lumbar puncture, as well as CIS patients that had developed into clinical definite MScl in the follow-up period, enabled the evaluation of the biomarker candidates as potential early diagnostic markers in MScl.

Stable Isotope Dilution-Selected Reaction Monitoring (SID-SRM) was applied for targeted verification of biomarker candidates, using stable isotope labeled internal standards (SISs) for targeted quantification. The SID-SRM-assays was developed by Ann Cathrine Kroksveen at PROBE. SID-SRM was used for verification of eight of the biomarker candidates from the discovery study; Alpha-1-acid glycoprotein 1, Alpha-2-macroglobulin, Ceruloplasmin, Complement C3, Prostaglandin H2 D-isomerase, Transthyretin, Cystatin C and Neuroserpin. In addition, SRM-assays were developed for a selection of proteins that had been reported in the literature with differential CSF abundance levels in MScl patients in previous proteomics-based biomarker discovery studies. The differential abundance of the literature-derived proteins had been reported in various group comparisons. In this study we wanted to investigate if there were an abundance difference for these proteins between novel patient and control cohorts. Furthermore, we wanted to investigate if these proteins could have potential as early diagnostic markers for MScl. Literature derived biomarker candidates included, Alpha-1-antichymotrypsin, Apolipoprotein D, Clusterin, Contactin-1, Kallikrein-6, Leucine-rich alpha-2-glycoprotein, Secretogranin-1 and Secretogranin-2, Serotransferrin and Serum albumin. Taken together, 18 biomarker candidates were included in the verification study of which the endogenous peptides were detectable in neat CSF. Supplementary Table 6.7 displays

the proteins and their signature peptides. Supplementary Table 6.8 lists biomarker candidates included in this study with literature references.

Signature peptides for each biomarker candidate protein were defined as previously explained (Section 2.4.2) and stable isotope labeled internal standards (SISs) corresponding to the signature peptides were purchased in crude quality from Thermo Scientific. Each CSF sample (n=125) comprising 10 µg protein was trypsin digested (Section 2.3.1) and identical amounts of SISs were spiked into each sample prior to desalting using C18 StageTips (Section 2.5.1). The amount of SIS used was similar to the amount of the endogenous form of the corresponding peptides determined in a CSF test sample used for SRM-assay development. Each sample was resuspended in 3% ACN/ 5% FA and 1 µg CSF protein digest was analyzed by LC SRM-MS on a Dionex Ultimate 3000RS nano-LC system coupled to a Q-Trap 5500 as described in Section 2.7.3. The peptides were analyzed by scheduled SRM, with the target scan time set to 1 second and the detection window 240 seconds. Three transitions per peptide were monitored, whereof one transition per peptide was selected for accurate relative quantification analysis. Supplementary Table 6.7 displays the signature peptides, transitions used for quantification, Q1 and Q3 m/z values for the endogenous and SIS peptides and the collision energy used in the SRM analysis.

Identical aliquots of a CSF test sample (i.e. pooled sample with CSF from multiple patients with various neurological conditions) were processed along side the patient CSF samples and spiked with SISs in the same way. These CSF test samples were used as quality controls of the sample processing and the analytical pipeline.

The SRM data was analyzed by use of the software MultiQuant 2.1.1 (AB SCIEX). Transition sets of SISs and endogenous peptides were automatically integrated followed by manual evaluation of the data whereof transitions displaying retention time errors, interference or poor integration were excluded from further analysis (Section 2.8.5). The most abundant transition for each target peptide was used for quantification. The fold change between the groups was calculated using the median SRM area ratio from each group (e.g. median area ratio MScl over median area ratio

control) and a Student t-test was used to calculate the p-value. A p-value less than 0.05 were considered significant.

3.3.2 SRM analysis of biomarker candidates from the discovery study

The significant decrease in abundance of Alpha-1-acid glycoprotein 1 found in the discovery experiment was verified by SRM as significant when comparing all MS patients (including CIS) and OIND controls. The fold change and respective p-value are listed in Table 3.3.1. Hence, comparison of the label-free and the SRM quantitative results indicates abundance decrease of this protein even in extended patient cohorts. Alpha-2-macroglobulin was significantly less abundant in CIS compared to OND. Cystatin C was the only protein with significant differential abundance in CIS compared to MScl. In addition, the CSF levels of Cystatin C was significant decreased comparing all MScl and OND, and CIS versus OND. Cystatin C and Neuroserpin were both showing significant increase in CSF level of MScl patients versus control in the discovery study, while in the verification study the differential expression of these proteins was decrease in CSF levels of CIS patients and OND. For Neuroserpin the significance was just outside the significance threshold (p-value of 0.0516).

3.3.3 SRM analysis of biomarker candidates selected from the literature

Significant abundance differences were found with SRM analysis for five of the ten proteins with reported differential abundance in MScl in the literature. Apolipoprotein D was significantly less abundant in RRMS versus OIND and OND. The CSF levels of Alpha-1-antichymotrypsin also were significant decreased in RRMS versus OIND, in addition to CIS compared with OIND. Combining all MScl in one comparison group gave the same significant decrease in abundance compared to OIND of the two proteins. Clusterin and Kallikrein-6 showed significant decrease in abundance level in all MScl when compared with OND. Kallikrein-6 also showed the same level of decrease in CIS versus OND. Tabel 3.3.3 and Figure 3.3.1 summarizes the results obtained in the biomarker verification study. Taken together significant differential abundance was found for seven of the 18 biomarker candidates using SID-SRM.

Comparison	Protein	Abundance	Fold change	P-value
Combined MScl vs. OIND	Alpha-1-acid glycoprotein	↓MScl	-1.1	0.0289
	Alpha-1-antichymotrypsin	↓MScl	-1.2	0.0021
	Apolipoprotein D	↓MScl	-1.2	0.0061
Combined MScl vs. OND	Apolipoprotein D	↓MScl	-1.2	0.0027
	Clusterin	↓MScl	-1.1	0.0499
	Kallikrein-6	↓MScl	-1.3	0.0179
	Cystatin C	↓MScl	-1.1	0.0196
RRMS vs OIND	Alpha-1-antichymotrypsin	↓RRMS	-1.5	0.0155
	Apolipoprotein D	↓RRMS	-1.2	0.0297
RRMS vs OND	Apolipoprotein D	↓RRMS	-1.2	0.0179
CIS-MS VS OND	Kallikrein-6	↓CIS-MS	-1.4	0.0049
CIS vs. OIND	Alpha-1-antichymotrypsin	↓CIS	-1.6	0.0023
CIS vs. OND	Alpha-2-macroglobulin	↓CIS	-1.3	0.0375
	Cystatin C	↓CIS	-1.6	0.0015
	Kallikrein-6	↓CIS	-1.5	0.0196
	Secretogranin-1	↓CIS	-1.3	0.0387
CIS vs. MScl	Cystatin C	↓CIS	-1.5	0.0137
OIND vs OND	Cystatin C	↓OIND	-1.1	0.0108
	Kallikrein-6	↓OIND	-1.4	0.0099
	Alpha-1-antichymotrypsin	↑OIND	1.6	0.0053

Table 3.3.1: Biomarker candidates showing significant abundance difference in the SID-SRM verification study. Fold change between compared groups was calculated from the median area ratio from SRM analysis and Student's t-test p-value less than 0.05 was considered significant.

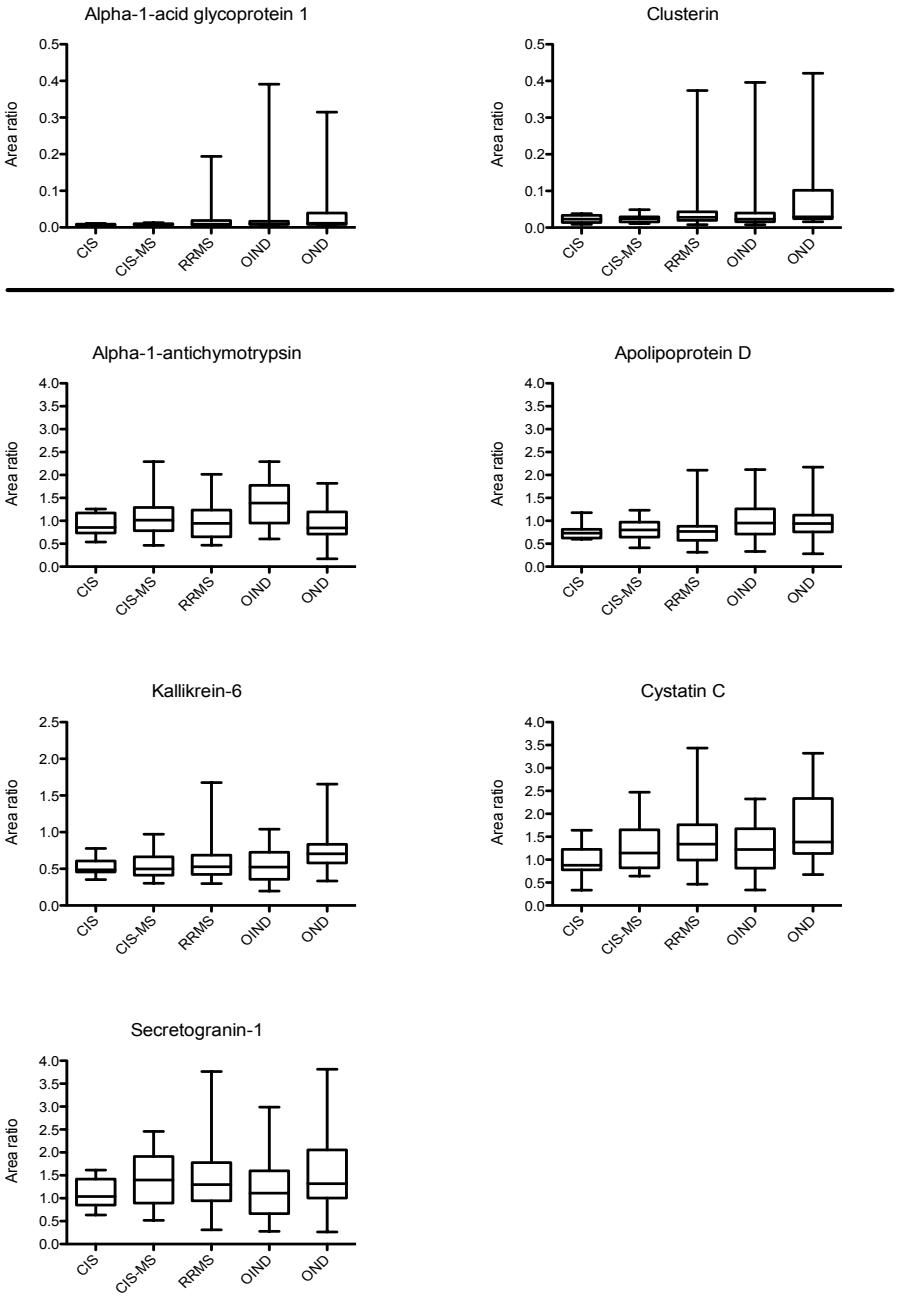

Figure 3.3.1: Proteins with significant differential CSF abundance in the SID-SRM analysis. Taken together, the presented proteins were in the SID-SRM analysis found to be differential abundant in several group comparisons. The respective group comparisons are shown in Table 3.3.1. The area ratios of the endogenous peptides used for quantification are plotted for each patient.

A protein is given the status of a biomarker candidate if its differential abundance is constant between two states in independent patient material using independent analytical methods. The most interesting results from the biomarker discovery and verification study was the significant differential abundance of Alpha-1-acid

glycoprotein 1 and Cystatin C derived from the initial screening phase, and the literature-derived candidate proteins Kallikrein-6 and Apolipoprotein D.

3.4 Characterization of the normal human CSF proteome

The schematic flow chart summarizes the main steps performed in the proteomics-based characterization of the normal CSF proteome and subsequent evaluation of size dependent differential abundance of CSF proteins in MScl. The focus of the mapping experiment was to perform extensive fractionation of the CSF proteome on protein level to obtain a comprehensive set of reference proteins that could further be used for investigations in MScl. We applied high abundant protein removal followed by separation of CSF proteins according to size using gradient SDS-PAGE. The size distribution of proteins across the gel for a selection of biomarker candidates in MScl was further examined. We observed the same protein present in several mass areas on the gel representing far different masses then what the mass of the full-length protein would imply. This observation indicated the presence of several protein variants for this protein. These protein variants could be truncation products, splice variants, or post transitionally modified proteins. This information is important to reveal and will serve as a first step on the way to individually quantify different protein variants originating from the same gene.

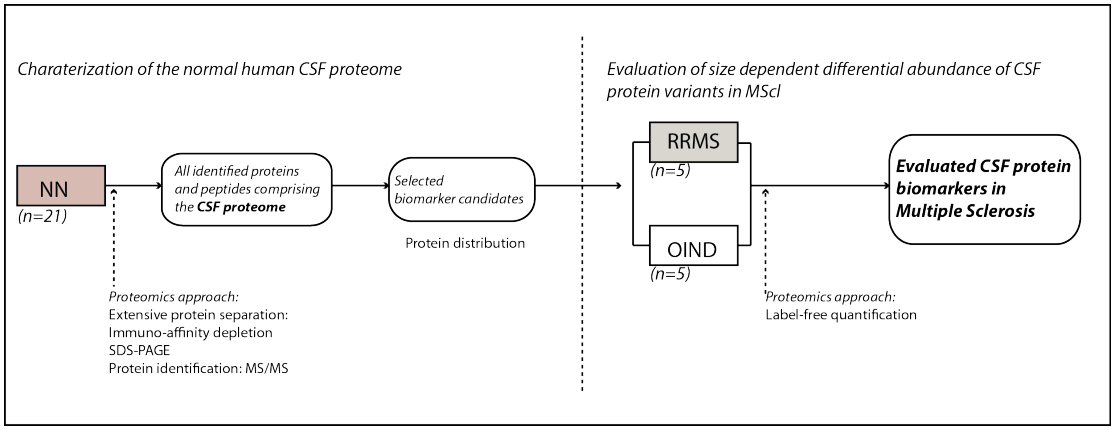

Figure 3.4.1: Characterization of the normal human CSF proteome and evaluation of biomarker protein variants in CSF of Mscl: The figure summarizes the different steps undertaken in characterization of the normal human CSF proteome and further evaluation of protein variants of biomarkers in CSF of MScl patients and controls. Characterization of the normal human CSF proteome was performed by use of CSF form non-neurological (NN) patients receiving spinal anesthesia prior minor surgery, to investigate size distribution of CSF proteins under normal conditions. The proteomics approach included extensive fractionation of the CSF proteome on protein level. We applied immuno-affinity depletion of 14 high abundant proteins in CSF followed by separating both the bound fraction and flow through fraction according to molecular mass using gradient SDS-PAGE. Based on the size distribution displayed by proteins in the normal CSF proteome, CSF from RRMS and OIND was subjected to evaluation of size dependent differential abundance of CSF protein variants. CSF from the two patient groups was separated according to size using SDS-PAGE. Following SDS-PAGE separation, the gel was divided giving rise to a high and a low molecular mass fraction that were analyzed using label-free relative quantification aiming to obtain quantitative information of the proteins differential abundance in RRMS in high and low molecular mass fractions.

3.4.1 Assessing blood contamination of the CSF samples

CSF could be contaminated with blood during lumbar puncture. In order to map the CSF proteome accurately, CSF samples contaminated with blood were avoided as these samples could introduce blood specific proteins to the CSF proteome. For the purpose of checking CSF for blood contamination a SID-SRM-assay was built with spike-in of stable isotope labeled internal standard (SIS) peptides representing hemoglobin, a blood specific protein. The hemoglobin SRM-assay was built together with Elise Aasebø (as described in Section 2.6).

Each CSF sample form non-neurological patient (n=50) was checked individually for blood contamination using the developed hemoglobin SRM-assay. CSF samples had already been trypsin digested and desalted by others at PROBE. The CSF samples containing 19 µg tryptic peptides were spiked with hemoglobin beta-1 SIS peptides at approximately the same amount as the corresponding endogenous peptide in the blood contaminated CSF test sample (Section 2.6). Each sample was analyzed by LC SRM-MS analysis on a Dionex Ultimate 3000RS nano-LC system coupled to a Q-Trap 4000.

The SRM data was analyzed using MultiQuant 2.1.1 (AB SCIEX). SISs and endogenous CSF hemoglobin beta-1 peptides were manually integrated and the data was evaluated in which transitions with wrong retention time, poor integrations and signal-to-noise less than 3 were considered to be free of signal from hemoglobin i.e. free of blood contamination. Of the 50 non-neurological CSF samples, 21 samples were considered as blood free as the hemoglobin levels were below the intensity threshold limit set for detection.

3.4.2 Experimental outline for mapping the normal CSF proteome

Each of the 21 samples without detectable blood contamination from non-neurological subjects contributed with 500 µL to the total of 10.5 mL pooled CSF. The protein concentration of the pool was measured to be 0.399µg/µL. 3.5 mL corresponding to an amount of approximately 1.40 mg protein was used in this experiment. The CSF pool was concentrated, purified and subjected to immuno-

affinity depletion of 14 high-abundant proteins (listed in Section 2.2.2). Immuno-affinity based partitioning generated two CSF protein fractions, a bound fraction consisting of the 14 high abundant proteins and their potential associated proteins (referred to as the bound fraction), and a flow-through fraction enriched with the less abundant proteins in CSF (referred to as the flow-through fraction in Figure 3.4.2). Both fractions were reduced and alkylated (Section 2.2.3) followed by protein separation according to molecular mass using a 5-15% gradient SDS-PAGE gel (Section 2.2.3).

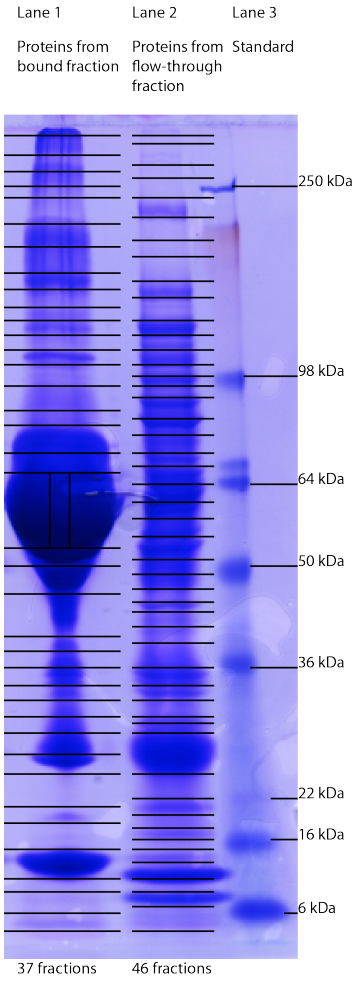

Figure 3.4.2: Separation of bound and flow-through fractions of the CSF proteome according to molecular mass. The CSF proteome was extensively fractionated on protein level. First line fractionation involved separating the proteome into two distinct fractions: the immuno-affinity depleted high abundant proteins, and the low abundant fraction (flow through). Both fractions were separated according to molecular mass using gradient SDS-PAGE. Lane 1 shows the bound fraction, while lane 2 displays the flow-through fraction. See Blue® Plus2 pre-stained protein standard was used as standard to get a certain impression of the size distribution during the separation, and is displayed in lane 3.

Albumin accounts for about 70 % of the total protein concentration in CSF, this is clearly visualized on the gel whereof a large amount of protein are found at 67 kDa.

Each of the excised bands in one of the two lanes on the gel represented a group of proteins at a specific molecular mass (Figure 3.4.2). A total of 83 fractions were generated, whereof 37 fractions were obtained from the depleted high abundant proteins and 46 from the flow through. Proteins were in gel digested using trypsin (Section 2.3.2) and peptides were desalted using C8 StageTips (Section 2.5.2). CSF protein digest from each individual fraction were resuspended in 0.1% FA prior to injection into the MS. A fraction of 0.5 μg was analyzed by LC-MS/MS for protein identification on an Orbitrap Velos Pro MS. The LC analysis was of 90 minutes duration and the top 15 most intense eluting peptides were sequentially isolated and fragmented in the high-pressure linear ion trap by CID (collision-induced-dissociation).

3.4.3 Protein and peptide identifications

The identification of proteins and peptides from the MS/MS data was achieved using both the Spectrum Mill search engine and the SearchGUI algorithm searching against the human SwissProt database with FDR of less than 1% as a filtering threshold (Section 2.8.1). The software Peptide Shaker was used in combination with SearchGUI to display the peptide and protein identifications. A total of 18, 807 distinct peptides were identified from all fractions, representing 1987 proteins (Protein and peptide information available in Supplementary Table 3.4 B). Among these 17, 824 peptides were identified in the low abundant fraction (flow through) (Lane 2, Figure 3.4.2) mapping to 1563 proteins. While 4001 peptides, mapping to 424 protein entities were identified in the high abundant fraction (Lane 1, Figure 3.4.2). The Venn diagram in Figure 3.4.3 compares the peptide identification obtained from the high and the low abundant portioning of the CSF proteome.

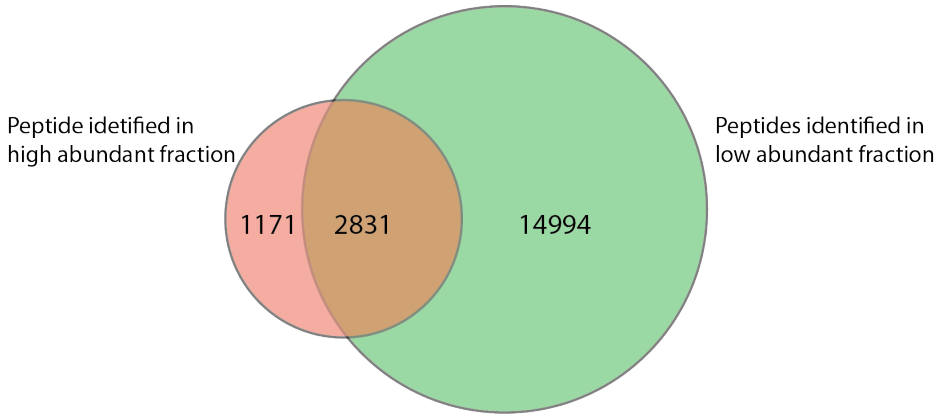

Figure 3.4.3: Venn diagram representation of all peptides identified in CSF of non-neurological subjects. Taken together, 18,807 peptides were identified. Extensive fractionation on protein level enabled a relative separation of low and high abundant CSF protein, leading to identification of 17,824 peptides in the low abundant fraction and a total of 4001 peptide identifications were obtained in the high abundant fraction. By aligning the identified peptides in the different fractions, the total of 14,994 peptides were exclusively identified among the low abundant proteins is CSF, whereas 1171 were exclusively identified as the high abundant fraction. Leading to 2831 peptides identified in both fractions.

The number of proteins identified in the two mass fractions on the gel is shown in Figure 3.4.4. The high number of proteins identified from the bound fraction are likely to be present in this fraction due to unspecific binding either to the column beads or to the 14 proteins targeted by the depletion column.

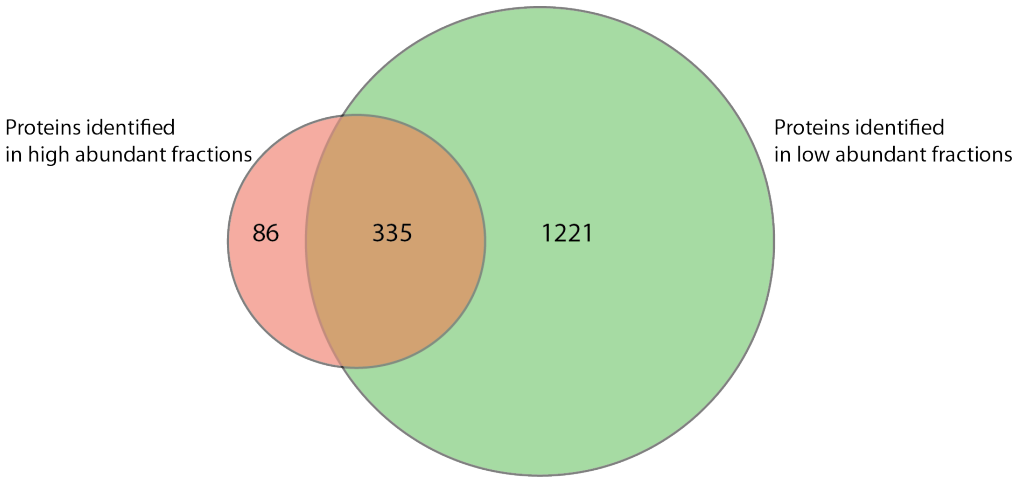

Figure 3.4.4: Proteins identified in the normal CSF proteome. By combining all raw data files in one search using SearchGUI, the total of 1987 proteins were identified. Of these 1563 were identified in the low abundant fraction and 424 in the high abundant fraction. Comparing the protein identifications obtained from the two gel fractions showed 86 protein entities representing the 14 proteins targeted by the depletion column. These 86 proteins were mostly isoforms of immunoglobulin.

3.4.4 Protein distribution in the CSF proteome

One of the aims in this experiment was to target a selection of biomarker candidates and investigate their size distribution on the SDS-PAGE gel to reveal potential protein variants (isoforms, truncation products and potential proteolytic processed products). The identification data generated from Spectrum Mill (Section 2.8.1) enabled protein distribution information to be obtained from the individual gel sections. Of special interest was the protein distribution of the low abundant portion of the CSF proteome. Each of the excised bands, or gel sections, represented a group of proteins at a specific molecular mass. Proteins were expected to be present in gel sections corresponding to their mass. As shown in figure 3.4.5 over 700 proteins were only identified in one gel section. A few proteins were distributed along all 46 gel sections; in which a general trend showed proteins with high intensities to be present in more than one fraction. Proteins of high abundance were typically present in all fractions.

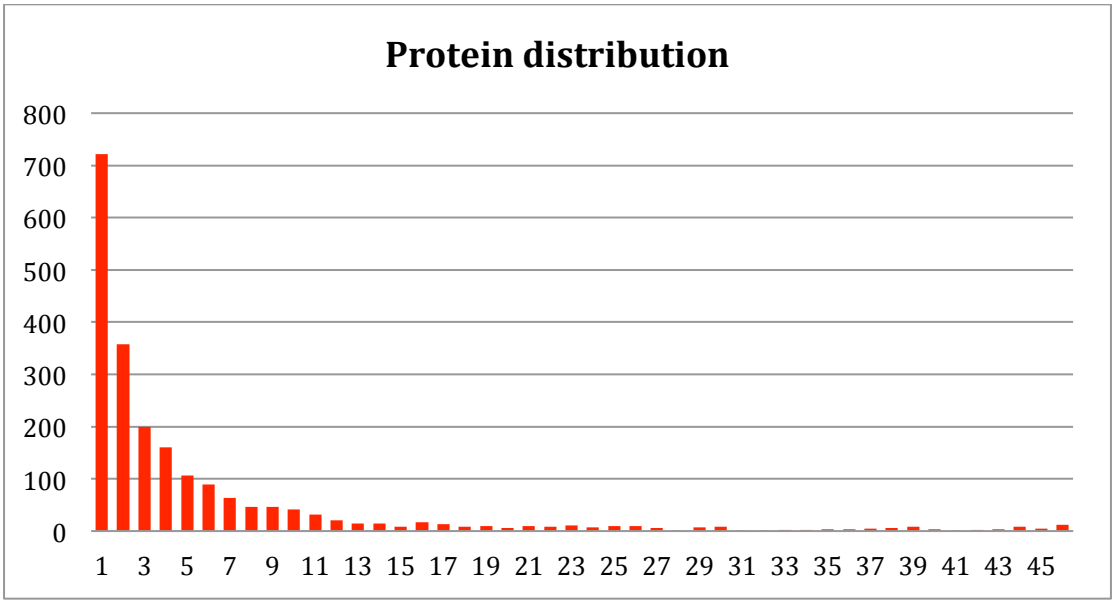

Figure 3.4.5: Protein distribution along the 46 fractions of low abundant proteins. The number of proteins is displayed on the x-axis, whereas the y-axis is displaying number of fractions. By processing each gel band as individual fractions, information of protein distribution was obtained. As showed in the distribution plot, over 700 proteins were observed in only one fraction. While a small number of proteins were identified in all 46 fractions.

A selection of biomarker candidates was chosen for further investigation based on their protein size distribution the 46 sections, in which the proteins was identified in mass areas on the gel representing far different masses then what the full-length protein mass would imply. The protein distribution of the biomarker candidates

Alpha-1-acid glycoprotein 1, Apolipoprotein D and Cystatin C are shown in Figure 3.4.6 A, B and C, respectively. We also chose to display the size distribution of Neuroserpin derived from the biomarker discovery study and the previously reported biomarker candidate Secretogranin-1 (Figure 3.4.6 D and E).

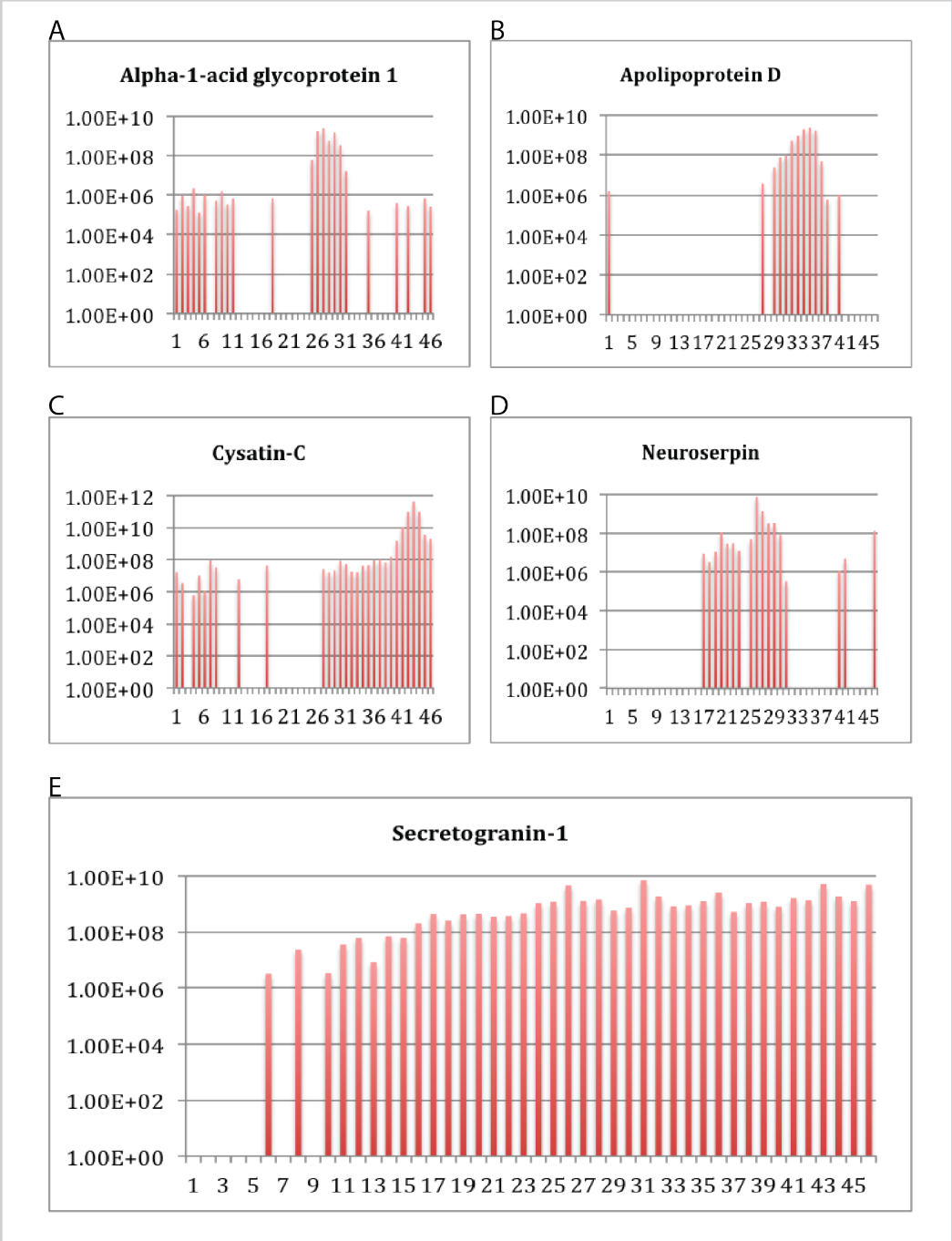

Figure 3.4.6: Distribution of a selection of biomarker candidates in the 46 gel sections.
By performing fractionation of the CSF proteome on protein level we were able to obtain size distribution information of a selection of biomarker candidates in MScl. The observation of the candidate protein present in several mass areas on the gel representing far different masses then what the mass of the full-length protein would imply indicated the presence of several protein variants for this protein.

3.4.5 Potential truncation products of the biomarker candidates

We hypothesized that the presence of truncation products of the proteins displayed in Figure 3.4.6 could be revealed by identifying non-tryptic peptides mapping to the proteins. By using the identification data generated from Spectrum Mill (Section 2.8.1) we could obtain information of the peptides identified for the proteins in the individual gel sections. The peptide identification data of Cystatin C revealed 13 non-tryptic peptides, whereas Secretogranin-1 revealed nine non-tryptic peptides. The presence of these peptides could further indicate the presence of several protein variants for each of these proteins in the normal CSF proteome. The identified non-tryptic peptides are shown in the full-length sequence of Cystatin C and Secretogranin-1 in Figure 3.4.7.

Cystatin C

MAGPLRAPLLLLAILAVALAVSPAAGSSPGKPPRLVGGPMDASVEEEGVRRALDFAVGEYNK
ASNDMYHSRALQVVRARKQIVAGVNYFLDVELGRTTCTKTQPNLDNCPFHDQPHLKRKAFC
SFQIYAVPWQGTMTLSKSTCQDA

Secretogranin-1

MQPTLLLSLLGAVGLAAVNSMPVDNRNHNEGMVTRCIIEVLSNALSKSSAPPITPECRQVLKT
SRKDVKDKETTENENTKFEVRLLRDPADASEAHESSSRGEAGAPGEEDIQGPTKADTEKWA
EGGGHSRERADEPQWSLYPSDSQVSEEVKTRHSEKSQREDEEEEEGENYQKGERGEDSS
EEKHLEEPGETQNAFLNERKQASAIKKEELVARSETHAAGHSQEKTHSREKSSQESGEETG
SQENHPQESKGQPRSQEESEEGEEDATSEVDKRRTRPRHHHGRSRPDRSSQGGSLPSEE
KGHPQEESEESNVSMASLGEKRDHHSTHYRASEEEPEYGEEIKGYPGVQAPEDLEWERYR
GRGSEEYRAPRPQSEESWDEEDKRNYPSLELDKMAHGYGEESEEERGLEPGKGRHHRGR
GGEPRAYFMSDTREEKRFLGEGHHRVQENQMDKARRHPQGAWKELDRNYLNYGEEGAP
GKWQQQGDLQDTKENREEARFQDKQYSSHHTAEKRKRLGELFNPYYDPLQWKSSHFERR
DNMNDNFLEGEEENELTLNEKNFFPEYNYDWWEKKPFSEDVNWGYEKRNLARVPKLDLKR
QYDRVAQLDQLLHYRKKSAEFPDFYDSEEPVSTHQEAENEKDRADQTVLTEDEKKELENLA
AMDLELQKIAEKFSQRG

Figure 3.4.7: Truncation products identified for the biomarker candidates Cystatin C and Secretogranin-1. By using the identification data generated from Spectrum Mill (Section 2.8.1) we could obtain information of non-tryptic peptides identified for the proteins in the individual gel sections. For Cystatin C 13 non-tryptic peptides were identified, while Secretogranin-1 revealed nine non-tryptic peptides.

3.5 Evaluation of size dependent differential abundance of CSF protein variants in Multiple Sclerosis

Based on the indications of presence of several protein variants for the biomarker candidate proteins in normal CSF, we wanted to investigate the findings in CSF of MScl. The aim of this experiment was to obtain quantitative information of protein abundance comparing high and low mass portions of the CSF proteome between RRMS and OIND. By separating CSF proteins of the two patient groups according to size using SDS-PAGE, the gel was divided into a high and a low mass fraction. Proteins present in both high and low mass areas on the gel were further targeted to obtain an abundance ratio between RRMS and controls and to investigate if these two ratios differed for the same protein.

3.5.1 Study design and analytical process

CSF from five RRMS and five OIND patients were included in the study, in which two pools were created, each comprising 10 µg CSF from five different patients in each group. Each CSF pool, containing a total of 50 µg protein, were reduced and alkylated (Section 2.2.3) and separated using a 4-12 % gradient NuPage® gel (Section 2.2.4).

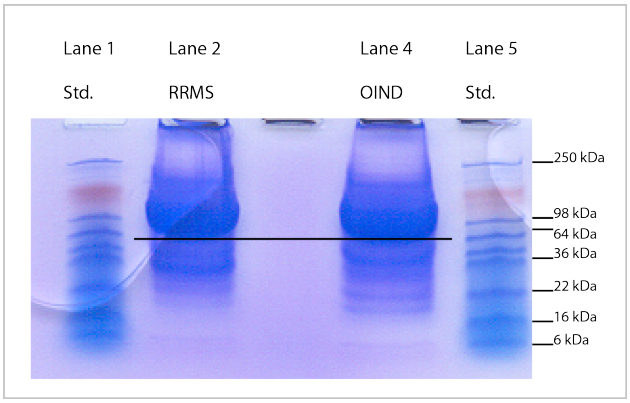

Figure 3.5.1: Separation of CSF proteins from RRMS and OIND patients. Lane 2 displays the separated CSF proteins of RRMS patients, while lane 4 displays the separated CSF proteins of OIND patients as controls. See Blue Plus2 pre-stained protein standard was used as standard (lane 1 and 5). The gel was cut at approximately 67 kDa under the band indicating the presence of serum albumin, resulting in a high and low mass fraction separation of CSF proteins from RRMS and OIND. The high and low mass fractions were further processed individually enabling the comparison of protein abundance in the two fractions between RRMS and OIND.

The gel was divided at 67 kDa resulting in a high and a low molecular mass separation of CSF proteins in RRMS and OIND. The four gel fractions were further processed individually. Each fraction was trypsin digestion (Section 2.3.2), desalted (Section 2.5.3), and dried samples were resuspended in 0.1 % FA. 1 µg CSF protein digest from each fraction was subjected to LC-MS/MS on a Dionex Ultimate 3000RS nano-LC system coupled online to an Orbitrap Velos Pro MS. The LC analysis was of 180 minute duration for extensive peptide separation.

The data resulting from the MS analysis were used for protein identification as well as label-free relative quantification analysis using Progenesis LC-MS (Section 2.8.2). By using this software the abundance of proteins identified in the high mass fraction of RRMS could be compared to the corresponding protein abundance in the high mass fraction of control. In the same manner protein abundance in low mass fractions could be compared across the two patient groups.

3.5.2 Protein identification

A total of 262 proteins were identified in the high mass fraction on the gel, while 292 proteins were present in the low mass fraction. By comparing the two mass fractions, 123 proteins were found to be present in both fractions (Figure 3.5.2).

Figure 3.5.2: Protein identifications in high and low molecular mass fractions. A total of 139 proteins were identified in the high molecular mass fraction, while 296 proteins were identified in the low molecular mass fraction. By comparing the protein identifications in the two mass fractions, 123 proteins were identified in both high and low mass areas on the gel.

3.5.3 Quantitative comparison of protein identified in high and low mass fractions in the CSF proteome of RRMS patients

A selection of the 123 proteins present in both mass areas on the gel were targeted for relative quantification between RRMS and controls, with the aim to determine if the ratio of the protein present in the two mass fractions differed for the same protein between the two patient groups. The selection of proteins to examine was based on the proteins showing interesting protein distribution (Figure 3.4.6).

Protein	Acc. number	Peptides used for quantification	Peptide count	Spectral count	Confidence score	High mass FC (RRMS/OIND)	Low mass FC (RRMS/OIND)
Serum albumin	P02768	114	114	1526	15801	1.0 ↑	1.1 ↑
Alpha-1-acid glycoprotein 1	P02763	8	11	52	1478	-1.3 ↓	-2.6 ↓
Apolipoprotein D	P05090	6	6	17	593	1.2 ↑	-1.3 ↓
Neuroserpin	Q99574	1	1	2	100	-1.5 ↓	-1.8 ↓
Cystatin-C	P01034	4	4	6	399	2.1 ↑	-1.5 ↓
Secretogranin-1	P05060	4	4	5	399	11.8 ↑	-1.5 ↓

Table 3.5.1: A selection of proteins identified in both high and low mass fractions. CSF from RRMS patients and OIND were separated according to size using SDS-PAGE and the proteins identified in both high and low mass areas on the gel was subjected to label-free relative quantification comparing protein abundance between the two patient groups. By using the software Progenesis LC-MS the fold change between RRMS and controls were obtained for proteins present in both high and low mass areas on the gel.

Serum albumin was found to be equally regulated in the two mass fractions. Similar regulation levels were also shown for Alpha-1-acid glycoprotein 1 and Neuroserpin being less abundant in RRMS compared to OIND in both mass fractions (Table 3.5.1). Cystatin C, Apolipoprotein D and Secretogranin-1 were among the oppositely regulated proteins. As displayed in Table 3.5.1, these proteins showed increased abundance in RRMS compared to OIND in the high mass fraction, while in the low mass fraction the proteins showed opposite regulation levels. The largest difference was observed for Secretogranin-1 that had an 11.8 fold increase in CSF levels in RRMS compared to OIND in the high mass fraction. While in the low mass fraction, the protein was decreased in RRMS (with a fold change of -1.5). Hence, Secretogranin-1 showed a 10 time higher intensity in the low mass fraction compared to the low intensity in the high mass fraction, as displayed in Figure 3.5.3. Distribution of Secretogranin-1 in the normal CSF proteome indicated the protein to be present in 40 fractions (Figure 3.4.6 E). The full-length protein of Secretogranin-1

has the molecular mass of 78 kDa, hence was expected to be present in fractions above 67 kDa.

	Peptide Quant	Peptides	Fold change (RRMS/OIND)	RRMS	Normalized abundance RRMS	Normalized abundance OIND
High mass fraction	4	SQREDEEEEEGENYQK HLEEPGETQNAFLNER GEAGAPGEEDIQGPTK CIIEVLSNALSK	11.8	↑	124937	10605
Low mass fraction	6	NYPSLELDK HLEEPGETQNAFLNER NYLNYGEEGAPGK CIIEVLSNALSK GEAGAPGEEDIQGPTK ASEEEPEYGEEIK	-1.5	↓	1231557	1864980

Table 3.5.2: CSF abundance of Secretogranin-1 in high and low mass fractions on the gel. The data obtained from label-free relative quantification approach using the software Progenesis LC-MS enabled the abundance of peptides identified in the high mass fraction of RRMS to be compared with the abundance of peptides mapping to Secretogranin-1 in the high mass fraction of OIND. The same comparison was performed for peptides identified in the low mass fraction. Shown in the table are peptide used for quantification of Secretogranin-1 in the two mass areas on the gel, fold changes between RRMS and OIND, regulation level in RRMS, and normalized protein abundance intensity are shown.

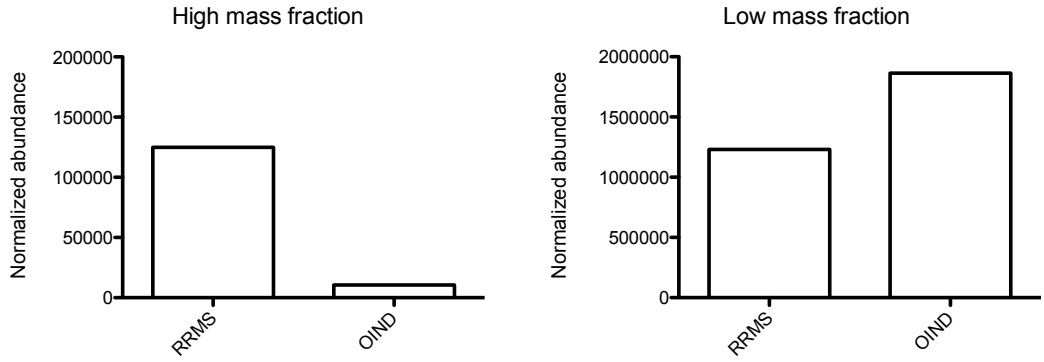

Figure 3.5.3: Secretogranin-1 show opposite regulation levels in high and low molecular mass fractions of CSF from RRMS patients. Normalized abundance levels of the Secretogranin-1 are shown. Secretogranin-1 showed a 10 time higher intensity in the low mass fraction compared to the high mass fraction.

Secretogranin-1 is a biomarker candidate previously been reported with differential abundance levels in MScl (Kroksveen 2012). Secretogranin-1 is a proprotein known to be proteolytically processed *in vivo* giving rise to biologically active peptide fragments (Benjannet, Leduc et al. 1987). The CSF concentrations of peptides derived from the protein have been reported as decreased in MScl (Mattsson, Ruetschi et al. 2007). The preliminary results of this study indicate differential CSF levels of

Secretogranin-1 full-length protein and protein variants in RRMS compared to OIND. Further studies could be done to investigate the involvement of peptide fragments derived from Secretogranin-1 in MScl pathogenesis.

4 Discussion

In this study we used mass spectrometry-based proteomics to study the cerebrospinal fluid proteome and revealed proteins involved in the pathogenesis of multiple sclerosis that could serve as biomarker candidates for the disease.

In the following sections, the quantification strategies used for biomarker discovery and verification are discussed in light of the observed abundance variations of the biomarker candidates across the different analytical approaches. Two main challenges in mass spectrometry-based proteomics are addressed: the complexity and dynamic range of protein composition in CSF, and the extent of biological variations. The four most interesting biomarker candidates from our studies are discussed in view of their pathological roles in MScl. Finally, the results obtained from the mapping experiment are discussed along with the further analysis that revealed opposite regulation level of the biomarker candidate Secretogranin-1 in high and low mass fractions of the CSF proteome. This section starts with a discussion around the patient and control groups.

4.1 Patient selection and controls

80-85 % of all MScl patients experience a relapsing-remitting disease course; hence biomarker discovery in RRMS patient represents a starting point for detecting significantly differential abundant proteins in clinical definite stages of the disease. Controls with OIND and OND are chosen over healthy controls. As MScl is considered to be an inflammatory disease involving neurodegenerative events, the use of other inflammatory neurological diseases (OIND) as controls minimize the risk of detecting general biomarkers for these pathological processes. The use of other neurological diseases (OND) represent an additional control layer, which further lowers the risk of identifying general markers for neurological diseases. In the SRM verification study we further included patients diagnosed with CIS at the time of lumbar puncture, as well as CIS patients that had developed into MScl, to enable the detection of abundance changes that take place at an early disease stage. Such proteins would potentially have great value as early diagnostic biomarkers for MScl.

4.2 Biomarker discovery in multiple sclerosis

In the biomarker discovery we applied a CSF pooling strategy and a label-free relative quantification approach to compare neat and depleted CSF proteomes between MScl patients and controls. The dynamic range in protein concentration in CSF is expected to be approximately nine orders of magnitude (Zhang 2007), while proteomics methods based on mass spectrometry have a detection limit within four or five orders of magnitude within a single mass spectrum (Rifai, Gillette et al. 2006). Hence, there is an imbalance between the complex CSF proteome and the MS instrumentation used for analysis, in which high abundant proteins are more likely to be detected, thus hampering the detection of lower abundant proteins. To reduce this problem, we chose high abundant protein depletion as a strategy to allow for broader proteome coverage. Each MScl and control pools were processed on two levels, enabling biomarker discovery in neat and depleted CSF (Section 3.1). The aim with the neat CSF strategy was to quantify high abundant proteins, while the depletion strategy was included to cover more of the proteome in the comparison, and this was indeed also observed from the result, with 60% more proteins identified after depletion. We discovered 65 proteins of differential abundance, in which only the 17 most promising were selected as biomarker candidates in MScl for further verification (Section 3.2 and 3.3). Some of the discovered biomarker candidates may be false positives; proteins that upon further evaluation are not differentially abundant. Due to limited number of patients included and the pooling strategy applied, and only three pools each of MScl and controls, the false discovery rate was expected to be rather high at this point of the biomarker pipeline. The pooling strategy was chosen to reduce the biological variance and at the same time keep the number of samples to be analyzed by MS relatively low.

4.3 The dimethyl-AIMS approach for confirmation of differential abundance

All phases that followed discovery in the biomarker pipeline replaced unbiased screening with targeted quantification approaches. We applied stable isotope dimethyl labeling coupled to the AIMS approach for targeted quantification of the 17 selected biomarker candidates from the screening phase. The dimethyl-AIMS approach was developed by Thin Thin Aye at PROBE, and as far as we know this is the first report

of this approach applied to biomarker qualification. We managed to identify 15 of the proteins in neat and individual CSF samples from 10 RRMS patients and 10 controls as described in Section 3.2. However we were not able to obtain an abundance ratio for all peptides in all experiments. This lack of quantification appears to be caused by the software used for analyzing the data, and could be improved if new software becomes available. The ten experiments could also have been analyzed by three technical replicates to improve detection and quantification confidence.

The biomarker qualification process resulted in 11 quantified proteins (Section 3.2). The only protein found to be significantly differential abundant was Transthyretin. As SID-SRM assays were already available for seven of the 11 quantified proteins, these proteins were also included in the SRM verification, even if not proven to have significant abundance changes from the qualification experiment. In the qualification study, the peptide abundance profiles obtained from these seven proteins were similar to that obtained from the screening phase, with exception of Complement C3 and Cystatin C contradicting the abundance levels of the two proteins. In the literature, increased abundance have been reported for Complement C3 (Stoop, Dekker et al. 2008), while decreased CSF abundance of Cystatin C have been reported in RRMS (Qin, Qin et al. 2009), supporting the results obtained from the dimethyl-AIMS analysis. The reason for not confirming more than one of the discovery findings as significant during the qualification study could be due to the limited number of samples included as discussed above in Section 4.2 and due to biological variations as discussed below in Section 4.5.

4.4 Selected reaction monitoring for biomarker verification

SID-SRM applied to biomarker verification allows for simultaneously targeting quantification of a set of peptides representing target proteins in larger patient cohorts. SID-SRM increases the reliability of quantification by enhancing both specificity and accuracy of analysis (Rifai, Gillette et al. 2006). We applied SID-SRM for targeted quantification of eight of the proteins discovered in the screening phase and ten literature derived candidate markers in a larger patient cohort (n=125) as described in Section 3.3. Due to the limited access of CSF available for fractionation, we chose to

focus our studies on high/medium abundant proteins in which endogenous peptides were detectable in neat CSF. The selection of literature markers and development of the SRM-assay was performed by Ann Cathrine Kroksveen, a researcher in the Multiple Sclerosis-group at PROBE. She also performed a separate SID-SRM analysis including 132 patients (125 used in the presented study), in addition to a few more proteins (Kroksveen 2012). By comparing the results, the significant decreased CSF level of Apolipoprotein D and Kallikrein-6 was further confirmed in CIS-MS and CIS patients compared with OND. Significant decreased abundance of Alpha-1-acid glycoprotein 1 was also found between RRMS patients and OND. SID-SRM analysis in 132 patients confirmed decreased abundance of Cystatin C in CIS patients compared to OND as found in the presented study. To our knowledge, this is the first report of a larger SRM verification study of biomarker candidates in MScl. There were some variations between the studies, and this could be due to not having quantitative data from all the same patients, as Kroksveen had slightly more patients in her study. In addition, not all proteins were quantified in all patients, and this also varied between the studies. In SRM analysis a protein is quantified based on signature peptides. The abundance of a signature peptide is not necessarily representative for the target protein as whole, significant abundance differences of a protein could appear depending on which peptide(s) that were used for quantification. This is probably due to the region of which the surrogate peptide represents could be potent for modifications, or representing a truncation or isoform of the protein (Lange, Picotti et al. 2008), therefore monitoring multiple peptides will increase the confidence of analysis. This could be one reason why not all literature candidates could be verified in this study as their initial quantification was based on different peptide sequences. A more likely explanation is that a low number of patients were used in the initial discovery experiment, and when we included more patients for verification, the results were no longer significant. The same explanation could be used for the biomarker candidates that we could not verify from our own discovery experiment.

4.5 The extent of disease and biological variation

In the verification phase of the biomarker pipeline the biological variation that was reduced in the screening phase by the pooling strategy was reintroduced. This was clearly demonstrated in the dimethyl-AIMS experiment, in which pronounced heterogeneity between the RRMS patients was observed. Biological variations and physiological processes such as the Circadian rhythm (Nilsson, Stahlberg et al. 1992), the rostro-caudal gradient, CSF flow-rate (Reiber 1994), and influx of plasma-derived proteins across the BBB, in addition to lifestyle, gender and age may influence the CSF protein composition. When screening CSF samples for differentially abundant proteins between MScl and controls, the pooling strategy reduced the biological heterogeneity. When pooling samples, it is important to recognize that if one patient included in a pool differs significantly from the rest, this one patient will affect the whole pool indicating an artificial abundance difference. As the regulation level for Complement C3 and Cystatin C observed in the qualification experiment contradicted the data obtained from the discovery, this could point towards the presence of outliers in one or more of the pools. 24 of the total 125 patients included in the extended SID-SRM verification study were also included in the initial screening phase. When analyzing these patients separately, although not significant, the same direction of regulation was observed as in the discovery experiment. Hence, the introduction of more patients in the verification study appeared to change the results for these proteins. The concentration of Serum albumin has been reported to vary from 0.097-0.403 g/L between CSF samples from neurologically normal individuals (Stoop, Coulier et al. 2010). Similar variations are typically significant for high abundant plasma proteins in CSF, such as Transthyretin and Haptoglobulin (Stoop, Coulier et al. 2010). Transthyretin was found to be significantly increased in CSF levels of MScl patients both in the screening phase (Section 3.1) as well as in the qualification experiment (Section 3.2), but not in the SID-SRM study. This could be explained by the reported high inter-individual variation of Transthyretin, which easily could lead to significant differences if a few patients are studied, as was the case for the discovery and qualification study. For CSF biomarker discovery it is essential to have an understanding of natural biological variation between individuals. Hence, CSF proteins with known high inter-individual variation should be assessed with caution as biomarker candidates in MScl. Taken together, we can conclude that increasing the

number of patients and controls in the screening phase will average out the biological variation and increase the statistical power of biomarker discovery experiments.

4.6 Promising biomarker candidates for multiple sclerosis

In the presented study we successfully identified proteins that may be involved in the pathology of MScl. The most interesting proteins from the presented study were Alpha-1-acid glycoprotein 1, Apolipoprotein D, Cystatin C, Kallikrein-6 and Secretogranin-1. The pathology of MScl is complex, involving both neuro-inflammatory and neurodegenerative processes.

4.6.1 Markers for immune engagement and inflammation

The first event in MScl pathology is expected to be of inflammatory character (Bielekova and Martin 2004) in which cellular inflammatory reactions are accompanied by disruption of the BBB and the entry of plasma proteins into the CNS (Lucchinetti, Bruck et al. 2001). The decreased abundance of Alpha-1-acid glycoprotein 1 found in the discovery experiment was confirmed by dimethyl-AIMS and SID-SRM analysis. Alpha-1-acid glycoprotein 1 is produced in response to inflammation and appears to function in modulating the activity of the immune system during acute phase reactions (Fournier, Medjoubi et al. 2000).

4.6.2 Markers for demyelination and neurodegeneration

Clearance of myelin debris from demyelinating axons may have important functional implications in MScl pathology. Apolipoprotein D, Cystatin C and Kallikrein-6 are involved in such processes. In the SRM verification study, the differential abundance of Apolipoprotein D was significant in several group comparisons (Section 3.3). Its abundance was significantly decreased in RRMS compared to OIND and OND, and in combining all MScl compared to the two control groups. Apolipoprotein D is involved in the removal of lipid during nerve degeneration and has previously been reported as a biomarker candidate in MScl (Reindl, Knipping et al. 2001). Decreased abundance of Apolipoprotein D have potential as marker for neurodegeneration and MScl due to its absence reported to cause decreased clearance of myelin during nerve

injury, resulting in delayed axonal regeneration and remyelination (Ganfornina, Do Carmo et al. 2010). Cystatin C is synthesized locally in the CNS and acts as a protease inhibitor of cathepsins. There is increased evidence of a possible role of this protein in the pathogenesis of MScl. After myelin destruction, myelin debris must be cleared by activated macrophages before the naked axons can be remyelinated. Activated macrophages produce the lysosomal protein cathepsins. Decreased abundance of Cystatin C and increased abundance of cathepsin B have been reported in CSF of MScl patients during a relapse (Nagai, Murakawa et al. 2000). We found Cystatin C to be of significantly decreased CSF levels in CIS compared to OND and all MScl compared to OND in the SID-SRM verification study. Kallikrein-6 has previously been reported with significantly differential abundance in MScl (Hammack, Fung et al. 2004). In the SRM verification Kallikrein-6 had significant decreased CSF levels in CIS and CIS-MS patients compared to OND. Furthermore, combining all MScl patients in one comparison group gave the same significant decrease. Kallikrein-6 is a serine protease known to degrade myelin protein (Scarisbrick, Blaber et al. 2002) and is found in elevated CSF levels at sites of inflammation in active MScl lesions. Markers for demyelination, remyelination and axonal damage may be more disease specific for MScl than inflammatory markers.

The four proteins discussed above have a function that fits well with the expected pathogenesis of MScl. It would be interesting to study these proteins further, regarding their role in the pathogenesis of MScl, and their diagnostic value by performing a larger verification study. No single protein found in our studies could clearly distinguish MScl patients from controls. Hence, their value as individual diagnostic markers appears to be limited, however they could be of value as part of a biomarker panel for MScl, where the different included proteins are reflecting different aspects of the disease and the sum have a diagnostic value.

4.7 Protein level fractionation and mapping of the normal CSF proteome enabled discovery of size dependent differential abundant protein variants in multiple sclerosis

The mapping experiment gave us the opportunity to examine the size distribution of biomarker candidate proteins in CSF (Section 3.4). The protein distribution across the gel of a selection of biomarker candidates was examined with the aim to reveal potential protein variants (isoforms, truncation products and proteolytic processed products). We observed proteins present in lower mass fractions than expected of their theoretical masses. We hypothesized that this could be due to enhanced proteolytic processing as CSF is naturally high in active proteases, or it could indicate the presence of unknown protein variants of the candidate proteins. The truncation theory was supported by a high number of non-tryptic peptides revealed for the candidate proteins Cystatin C and Secretogranin-1.

The experiment had immediate utility for investigations in MScl. Based on our observations in normal CSF, we applied a label-free relative quantification approach aiming to obtain an abundance ratio between RRMS and controls of proteins present in high and low mass areas of the gel and to investigate if these ratios differed for the same proteins. The most striking observation was of Secreogranin-1, which showed an opposite CSF regulation level in RRMS comparing the abundance in the two mass fractions. The Multiple Sclerosis-group at PROBE has previously reported Secretogranin-1 as a promising early diagnostic biomarker candidate, in which the protein showed significant decrease in RRMS compared to OIND. Further, it was found to be significantly increased in CIS-MS compared to RRMS (Kroksveen 2012). Decreased abundance of two Secretogranin-1 peptides has been found in patients with clinical definite MScl (Mattsson, Ruetschi et al. 2007). These findings could indicate the abundance of Secretogranin-1 to decrease during the disease course of MScl. Our findings indicates that the full-length protein present in the high mass fraction is highly increased in MScl, while protein variants of lower mass, potentially being truncation products, are decreased in CSF levels in RRMS patients compared to control. Its observed regulation in MScl and its biological role in synaptic vessel (Taupenot, Harper et al. 2003) makes Secretogranin-1 an interesting biomarker candidate which should be further analyzed. Further adding to its significance,

Secretogranin-1 is a proprotein known to be proteolytic processed *in vivo* giving rise to biologically active peptide fragments (Benjannet, Leduc et al. 1987). Such peptides could reflect biological events and contain diagnostic biomarkers due to the fact that in many degenerative diseases there is an abnormal enzymatic degradation of proteins, thus rendering the presence of circulating peptides as disease specific and as a consequence of diagnostic value. The preliminary results obtained from this pilot study show highly interesting indications, however further analysis needs to be performed in order to determine the possible significance of Secretogranin-1 derived peptide fragments in MScl.

Future perspectives

In respect to the biomarker discovery, a new discovery experiment should be perform including more patients to average out the heterogeneity between samples and increase the quantification confidence. Extensive fractionation on both protein and peptide level could be applied for increased CSF proteome coverage.

SRM analysis should be extended to include more patients and controls to determine the clinical utility of the biomarker candidates. Furthermore, more peptides per protein should be included for increasing the confidence of quantification. The CSF proteome mapping experiment could serve as a reference library of detectable peptides and proteins.

It would be of great interest to investigate the regulation level of Secretogranin-1 peptides in MScl pathogenesis. We identified a total of 72 peptides mapping to Secretogranin-1. These peptides could serve as signature peptide for Secretogranin-1 protein variants for SID-SRM analysis.

In addition, as CSF naturally contains high concentrations of proteases, the neurodegenerative aspects regarding abnormal enzymatic degradation of proteins in CSF of MScl patients could be addressed by using protease inhibitors to investigate if the CSF peptides results from *in vivo* processing or due to sample treatment prior MS analysis.

5 References

Aebersold, R. and M. Mann (2003). "Mass spectrometry-based proteomics." Nature **422**(6928): 198-207.

Bantscheff, M., M. Schirle, et al. (2007). "Quantitative mass spectrometry in proteomics: a critical review." Anal Bioanal Chem **389**(4): 1017-1031.

Benjannet, S., R. Leduc, et al. (1987). "Chromogranin B (secretogranin I), a putative precursor of two novel pituitary peptides through processing at paired basic residues." FEBS Lett **224**(1): 142-148.

Bielekova, B. and R. Martin (2004). "Development of biomarkers in multiple sclerosis." Brain **127**(Pt 7): 1463-1478.

Blennow, K., P. Fredman, et al. (1993). "Protein analyses in cerebrospinal fluid. I. Influence of concentration gradients for proteins on cerebrospinal fluid/serum albumin ratio." Eur Neurol **33**(2): 126-128.

Boersema, P. J., T. T. Aye, et al. (2008). "Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates." Proteomics **8**(22): 4624-4632.

Bruck, W. and C. Stadelmann (2003). "Inflammation and degeneration in multiple sclerosis." Neurol Sci **24 Suppl 5**: S265-267.

Comabella, M., M. Fernandez, et al. (2010). "Cerebrospinal fluid chitinase 3-like 1 levels are associated with conversion to multiple sclerosis." Brain **133**(Pt 4): 1082-1093.

Compston, A. and A. Coles (2002). "Multiple sclerosis." Lancet **359**(9313): 1221-1231.

Compston, A. and A. Coles (2008). "Multiple sclerosis." Lancet **372**(9648): 1502-1517.

Dalgleish, A. G. (1997). "Viruses and multiple sclerosis." Acta Neurol Scand Suppl **169**: 8-15.

de Hoog, C. L. and M. Mann (2004). "Proteomics." Annu Rev Genomics Hum Genet **5**: 267-293.

Ferguson, B., M. K. Matyszak, et al. (1997). "Axonal damage in acute multiple sclerosis lesions." Brain **120 (Pt 3)**: 393-399.

Fournier, T., N. N. Medjoubi, et al. (2000). "Alpha-1-acid glycoprotein." Biochim Biophys Acta **1482**(1-2): 157-171.

Gallien, S., E. Duriez, et al. (2011). "Selected reaction monitoring applied to proteomics." J Mass Spectrom **46**(3): 298-312.

Ganfornina, M. D., S. Do Carmo, et al. (2010). "ApoD, a glia-derived apolipoprotein, is required for peripheral nerve functional integrity and a timely response to injury." Glia **58**(11): 1320-1334.

Gevaert, K. and J. Vandekerckhove (2000). "Protein identification methods in proteomics." Electrophoresis **21**(6): 1145-1154.

Grytten, N., S. B. Glad, et al. (2006). "A 50-year follow-up of the incidence of multiple sclerosis in Hordaland County, Norway." Neurology **66**(2): 182-186.

Hammack, B. N., K. Y. Fung, et al. (2004). "Proteomic analysis of multiple sclerosis cerebrospinal fluid." Mult Scler **10**(3): 245-260.

Han, X., A. Aslanian, et al. (2008). "Mass spectrometry for proteomics." Curr Opin Chem Biol **12**(5): 483-490.

Harris, V. K. and S. A. Sadiq (2009). "Disease biomarkers in multiple sclerosis: potential for use in therapeutic decision making." Mol Diagn Ther **13**(4): 225-244.

Hsu, J. L., S. Y. Huang, et al. (2003). "Stable-isotope dimethyl labeling for quantitative proteomics." Anal Chem **75**(24): 6843-6852.

Jacobs, L. D., R. W. Beck, et al. (2000). "Intramuscular interferon beta-1a therapy initiated during a first demyelinating event in multiple sclerosis. CHAMPS Study Group." N Engl J Med **343**(13): 898-904.

Jaffe, J. D., H. Keshishian, et al. (2008). "Accurate inclusion mass screening: a bridge from unbiased discovery to targeted assay development for biomarker verification." Mol Cell Proteomics **7**(10): 1952-1962.

Keegan, B. M. and J. H. Noseworthy (2002). "Multiple sclerosis." Annu Rev Med **53**: 285-302.

Kirkpatrick, D. S., S. A. Gerber, et al. (2005). "The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications." Methods **35**(3): 265-273.

Kroksveen, A. C. (2012). Quantitative proteomics analysis of cerebrospinal fluid.

Lange, V., P. Picotti, et al. (2008). "Selected reaction monitoring for quantitative proteomics: a tutorial." Mol Syst Biol **4**: 222.

Lassmann, H. (1998). "Neuropathology in multiple sclerosis: new concepts." Mult Scler **4**(3): 93-98.

Lucchinetti, C., W. Bruck, et al. (2001). "Multiple sclerosis: recent developments in neuropathology, pathogenesis, magnetic resonance imaging studies and treatment." Curr Opin Neurol **14**(3): 259-269.

Lucchinetti, C., W. Bruck, et al. (2000). "Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination." Ann Neurol **47**(6): 707-717.

Mallick, P. and B. Kuster (2010). "Proteomics: a pragmatic perspective." Nat Biotechnol **28**(7): 695-709.

Mattsson, N., U. Ruetschi, et al. (2007). "Cerebrospinal fluid concentrations of peptides derived from chromogranin B and secretogranin II are decreased in multiple sclerosis." J Neurochem **103**(5): 1932-1939.

McDonald, W. I., A. Compston, et al. (2001). "Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis." Ann Neurol **50**(1): 121-127.

Miller, D., F. Barkhof, et al. (2005). "Clinically isolated syndromes suggestive of multiple sclerosis, part I: natural history, pathogenesis, diagnosis, and prognosis." Lancet Neurol **4**(5): 281-288.

Miller, D. H., D. T. Chard, et al. (2012). "Clinically isolated syndromes." Lancet Neurol **11**(2): 157-169.

Nagai, A., Y. Murakawa, et al. (2000). "Cystatin C and cathepsin B in CSF from patients with inflammatory neurologic diseases." Neurology **55**(12): 1828-1832.

Nesvizhskii, A. I., O. Vitek, et al. (2007). "Analysis and validation of proteomic data generated by tandem mass spectrometry." Nat Methods **4**(10): 787-797.

Nilsson, C., F. Stahlberg, et al. (1992). "Circadian variation in human cerebrospinal fluid production measured by magnetic resonance imaging." Am J Physiol **262**(1 Pt 2): R20-24.

Noseworthy, J. H., C. Lucchinetti, et al. (2000). "Multiple sclerosis." N Engl J Med **343**(13): 938-952.

Ottervald, J., B. Franzen, et al. (2010). "Multiple sclerosis: Identification and clinical evaluation of novel CSF biomarkers." J Proteomics **73**(6): 1117-1132.

Polman, C. H., S. C. Reingold, et al. (2005). "Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria"." Ann Neurol **58**(6): 840-846.

Qin, Z., Y. Qin, et al. (2009). "Alteration of DBP levels in CSF of patients with MS by proteomics analysis." Cell Mol Neurobiol **29**(2): 203-210.

Regeniter, A., J. Kuhle, et al. (2009). "A modern approach to CSF analysis: pathophysiology, clinical application, proof of concept and laboratory reporting." Clin Neurol Neurosurg **111**(4): 313-318.

Reiber, H. (1994). "Flow rate of cerebrospinal fluid (CSF)--a concept common to normal blood-CSF barrier function and to dysfunction in neurological diseases." J Neurol Sci **122**(2): 189-203.

Reiber, H. (2001). "Dynamics of brain-derived proteins in cerebrospinal fluid." Clin Chim Acta **310**(2): 173-186.

Reindl, M., G. Knipping, et al. (2001). "Increased intrathecal production of apolipoprotein D in multiple sclerosis." J Neuroimmunol **119**(2): 327-332.

Rejdak, K., S. Jackson, et al. (2010). "Multiple sclerosis: a practical overview for clinicians." Br Med Bull **95**: 79-104.

Rifai, N., M. A. Gillette, et al. (2006). "Protein biomarker discovery and validation: the long and uncertain path to clinical utility." Nat Biotechnol **24**(8): 971-983.

Rithidech, K. N., L. Honikel, et al. (2009). "Protein expression profiles in pediatric multiple sclerosis: potential biomarkers." Mult Scler **15**(4): 455-464.

Sadovnick, A. D., G. C. Ebers, et al. (1996). "Evidence for genetic basis of multiple sclerosis. The Canadian Collaborative Study Group." Lancet **347**(9017): 1728-1730.

Scarisbrick, I. A., S. I. Blaber, et al. (2002). "Activity of a newly identified serine protease in CNS demyelination." Brain **125**(Pt 6): 1283-1296.

Schirle, M., M. A. Heurtier, et al. (2003). "Profiling core proteomes of human cell lines by one-dimensional PAGE and liquid chromatography-tandem mass spectrometry." Mol Cell Proteomics **2**(12): 1297-1305.

Segal, M. B. (1993). "Extracellular and cerebrospinal fluids." J Inherit Metab Dis **16**(4): 617-638.

Smith, D. E., C. E. Johanson, et al. (2004). "Peptide and peptide analog transport systems at the blood-CSF barrier." Adv Drug Deliv Rev **56**(12): 1765-1791.

Stoop, M. P., L. Coulier, et al. (2010). "Quantitative proteomics and metabolomics analysis of normal human cerebrospinal fluid samples." Mol Cell Proteomics **9**(9): 2063-2075.

Stoop, M. P., L. J. Dekker, et al. (2008). "Multiple sclerosis-related proteins identified in cerebrospinal fluid by advanced mass spectrometry." Proteomics **8**(8): 1576-1585.

Taupenot, L., K. L. Harper, et al. (2003). "The chromogranin-secretogranin family." N Engl J Med **348**(12): 1134-1149.

Teunissen, C. E., A. Petzold, et al. (2009). "A consensus protocol for the standardization of cerebrospinal fluid collection and biobanking." Neurology **73**(22): 1914-1922.

Trapp, B. D., L. Bo, et al. (1999). "Pathogenesis of tissue injury in MS lesions." J Neuroimmunol **98**(1): 49-56.

Trapp, B. D. and K. A. Nave (2008). "Multiple sclerosis: an immune or neurodegenerative disorder?" Annu Rev Neurosci **31**: 247-269.

Tumani, H., H. P. Hartung, et al. (2009). "Cerebrospinal fluid biomarkers in multiple sclerosis." Neurobiol Dis **35**(2): 117-127.

Vaudel, M., H. Barsnes, et al. (2011). "SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches." Proteomics **11**(5): 996-999.

Washburn, M. P., D. Wolters, et al. (2001). "Large-scale analysis of the yeast proteome by multidimensional protein identification technology." Nat Biotechnol **19**(3): 242-247.

Wilkins, M. R., C. Pasquali, et al. (1996). "From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis." Biotechnology (N Y) **14**(1): 61-65.

You, J. S., V. Gelfanova, et al. (2005). "The impact of blood contamination on the proteome of cerebrospinal fluid." Proteomics **5**(1): 290-296.

Zhang, J. (2007). "Proteomics of human cerebrospinal fluid - the good, the bad, and the ugly." Proteomics Clin Appl **1**(8): 805-819.

6 Appendix

Due to the size and format an Additional appendix (Supplementary Tables 3.1B-3.5B) is available at a local computer (EIR server) at PROBE.

6.1 Patient information for patients included in the biomarker discovery experiment

Pool	Diagnosis	Gender	Age at LP	Protein conc. (µg/mL)	OCB
MScl 1	RRMS	F	47	638	Positive
	RRMS	M	31	502	Positive
	RRMS	F	38	411	Positive
	RRMS	F	28	442	Positive
	RRMS	F	51	404	Positive
MScl 2	RRMS	F	35	325	Positive
	RRMS	F	45	235	Positive
	RRMS	F	33	396	Positive
	RRMS	M	59	825	Positive
	RRMS	F	45	598	Positive
MScl 3	CIS (MS)	M	43	339	Positive
	PPMS	F	41	511	Positive
	RRMS	F	35	229	Positive
	RRMS	F	42	361	Positive
	RRMS	F	39	266	Positive
	RRMS	F	46	438	Positive
OIND 1	Meningitis	F	31	1370	Negative
	Bells palsy	F	35	834	Negative
	GBS	M	53	1140	Negative
	GBS	M	32	953	Negative
	Meningoradikulitt	M	35	351	Negative
OIND 2	PML	F	89	755	Negative
	Neuroborreliosis	F	56	426	Negative
	Neuroborreliosis	M	19	433	Negative
	Rombencephalitt	M	18	344	Positive
	Borreliosis	F	59	374	Negative
OND 3	Pain	M	71	366	Negative
	ALS	F	58	399	Negative
	Asthenia	M	27	425	Negative
	Small-vessel-disease	M	62	526	Negative
	Parkinsonism	M	69	668	Negative

6.2 Patient information for patients included in the biomarker qualification experiment

Patient	Diagnosis	Gender	Age at LP	Protein conc. (µg/mL)	OCB
MScl	RRMS	F	28	411	Positive
MScl	RRMS	F	32	394	Positive
MScl	RRMS	F	43	291	Positive
MScl	RRMS	F	43	469	Positive
MScl	RRMS	F	28	490	Positive
MScl	RRMS	F	23	477	Positive
MScl	RRMS	F	37	508	Positive
MScl	RRMS	F	64	378	Positive
MScl	RRMS	F	42	361	Positive
MScl	RRMS	F	46	438	Positive
OIND	Latent neurosyfilis	F	28	623	Negative
OIND	Zoster nevralgi	F	58	707	Positive
OIND	Nevroboreliosis	F	45	590	Positive
OIND	Bells parasis	M	26	713	Positive
OIND	PML	F	89	755	Positive
OIND	PML	M	35	529	Negative
OIND	Parestestier	F	47	376	Negative
OIND	Vasculitt	F	27	482	Negative
OIND	Rombencephalitt	M	18	344	Positive
OIND	Boreliosis	F	59	374	Negative

6.3 Patient information for patients included in the biomarker verification

Patient	Gender	Age at LP	Protein conc. (µg/mL)	OCB
CIS	F	47	355	Positive
CIS	F	37	313	Positive
CIS	F	36	423	Positive
CIS	F	25	427	Positive
CIS	F	17	328	Positive
CIS	F	25	436	Positive
CIS	F	33	477	Positive
CIS	M	38	365	Positive
CIS	M	22	372	Positive
CIS	M	27	422	Positive
CIS	F	27	496	Positive
CIS	F	53	481	Positive
CIS	F	19	355	Positive
CIS-MS	F	26	490	Positive
CIS-MS	F	34	512	Positive
CIS-MS	F	20	426	Positive
CIS-MS	M	59	773	Positive
CIS-MS	M	43	551	Positive
CIS-MS	F	37	338	Positive
CIS-MS	F	34	430	Positive
CIS-MS	F	26	385	Positive
CIS-MS	M	33	412	Positive
CIS-MS	M	33	546	Positive
CIS-MS	F	19	341	Positive
CIS-MS	F	30	340	Positive
CIS-MS	F	52	358	Positive
CIS-MS	F	30	986	Positive
RRMS	M	45	522	Positive
RRMS	F	37	455	Positive
RRMS	M	24	494	Negative
RRMS	M	38	482	Positive
RRMS	F	26	543	Positive
RRMS	F	59	426	Positive
RRMS	F	36	336	Positive
RRMS	M	33	578	Positive
RRMS	F	42	373	Positive
RRMS	F	26	525	Positive
RRMS	F	28	399	Positive
RRMS	F	48	562	Positive
RRMS	F	30	544	Positive
RRMS	F	31	502	Positive
RRMS	F	38	472	Positive
RRMS	F	28	442	Positive
RRMS	F	28	411	Positive
RRMS	F	51	404	Positive
RRMS	F	35	338	Positive
RRMS	F	39	235	Positive
RRMS	F	33	396	Positive
RRMS	F	32	394	Positive

RRMS	F	29	294	Positive
RRMS	F	43	291	Positive
RRMS	F	43	469	Positive
RRMS	F	28	490	Positive
RRMS	F	23	477	Positive
RRMS	F	37	508	Positive
RRMS	F	45	598	Positive
RRMS	M	29	339	Positive
RRMS	F	35	229	Positive
RRMS	F	64	378	Positive
RRMS	F	42	361	Positive
RRMS	F	39	266	Positive
RRMS	F	45	438	Positive
OIND	F	41	897	Negative
OIND	F	48	423	No info.
OIND	M	71	502	Negative
OIND	F	60	462	Negative
OIND	F	39	426	Positive
OIND	M	54	400	Negative
OIND	F	45	345	Positive
OIND	F	40	621	Negative
OIND	F	26	263	Negative
OIND	M	43	621	Positive
OIND	M	42	922	Negative
OIND	M	59	358	Negative
OIND	F	52	354	Negative
OIND	F	62	619	Negative
OIND	F	40	385	Negative
OIND	F	31	1370	Negative
OIND	F	35	834	Negative
OIND	F	45	623	Negative
OIND	F	58	707	Negative
OIND	M	53	1140	Negative
OIND	M	26	713	Positive
OIND	F	65	576	Negative
OIND	M	31	953	Negative
OIND	M	35	351	Negative
OIND	F	89	755	Negative
OIND	M	35	529	Positive
OIND	F	47	376	Negative
OIND	F	27	482	Negative
OIND	F	63	426	Negative
OIND	M	43	433	Negative
OIND	M	18	344	Positive
OIND	F	59	374	Negative
OND	M	62	435	Negative
OND	F	42	266	Negative
OND	F	66	637	Negative
OND	M	72	439	Negative
OND	F	73	592	Negative
OND	M	63	674	Negative
OND	M	59	467	Negative
OND	F	51	505	Negative
OND	F	38	378	Negative
OND	F	43	325	Negative
OND	F	50	354	Negative

OND	F	53	569	Negative
OND	M	24	325	Negative
OND	F	52	403	Negative
OND	M	41	654	Negative
OND	M	68	542	Negative
OND	M	45	442	Negative
OND	M	61	631	Negative
OND	F	32	336	Negative
OND	F	64	314	Negative
OND	F	58	399	Negative
OND	M	27	425	Negative
OND	F	54	520	Negative
OND	M	69	668	Negative
OND	F	31	303	Negative
OND	M	55	566	Negative
OND	M	51	378	Negative
OND	M	43	496	Negative
OND	M	40	372	Negative
OND	M	71	366	Negative
OND	M	62	526	Negative

*OIND and OND was diagnosed with similar categories of diseases as in Supplemetary Table 6.1 and 6.2.

6.4 Patient information of non-neurological patients included for characterization of the normal human CSF proteome

Patient	Gender	Age at LP	Protein conc. (µg/mL)
NN	F	87	280
NN	M	61	376
NN	M	36	562
NN	M	80	629
NN	M	79	550
NN	F	82	555
NN	M	60	666
NN	M	71	590
NN	F	64	314
NN	M	68	584
NN	M	19	276
NN	M	58	599
NN	M	39	531
NN	F	74	521
NN	M	25	353
NN	F	28	379
NN	F	40	289
NN	M	63	511
NN	F	86	496
NN	F	45	281
NN	M	19	407

6.5 Patient information of patients included in the evaluation of size dependent differential abundance of CSF protein variants in multiple sclerosis

Patient	Gender	Age at LP	Protein conc. (µg/mL)
RRMS	F	43	469
RRMS	F	28	490
RRMS	F	22	476
RRMS	F	36	508
RRMS	F	45	438
OIND	F	27	482
OIND	F	28	623
OIND	M	26	713
OIND	F	47	375
OIND	F	89	755

6.6 Batch number for SISs ordered for the respective target proteins for SID-SRM assays

Supplier	Protein	Batch number
Thermo Scientific	Neuroserpin	OR16024
	Alpha-1-acid glycoprotein 1	SO303458
	Alpha-2-macroglobulin	OR284584
	Ceruloplasmin	SO303458
	Complement C3	SO303458
	Transthyretin	OR284584
	Cystatin C	SO303458
	Alpha-1-antichymotrypsin	OR16024
	Apolipoprotein D	OR286250
	Clusterin	OR284584
	Contactin-1	OR286250
	Kallikrein-6	OR284584
	Leucine-rich alpha-2-glycoprotein	OR16024
	Secretogranin-1	SO303458
	Secretogranin-2	SO303458
	Serotransferrin	SO303458
	Serum albumin	SO303458
	Hemoglobin beta-1	OR284584

6.7 Proteins and peptides used for SRM verification

Protein ID.	Acc. no	Peptide used in SRM	Transition used for quantification	Endogenous peptide		SIS peptide		CE
				Q1	Q3	Q1	Q3	
Alpha-1-acid glycoprotein	P02763	TEDTIFLR	y6	498	764	503.0	774.0	48
Alpha-1-antichymotrypsin	P01011	EQLSLLDR	y5	487	603	492.0	613.0	38
Alpha-2-macroglobulin	P01023	AIGYLNTGYQR	y6	628.2	738.3	633.2	748.3	37
Neuroserpin	Q99574	QEVPLATLEPLVK	y10	718.9	1080.6	722.9	1088.6	37
Apolipoprotein D	P05090	NILTSNNIDVK	y8	615.9	890.4	619.9	898.4	46
Ceruloplasmin	P00450	GAYPLSIEPIGVR	y8	686.8	870.5	691.8	880.5	46
Clusterin	P10909	IDSLLENDR	y5	537.6	646.3	542.6	656.3	35
Complement C3	P01024	ISLPESLK	y5	444	573.3	448	581.3	32
Secretogranin 1		GEAGAPGEEDIQGPTK	y11	779	1170	783	1178	39
Contactin-1	Q12860	DGEYVVEVR	y6	533.4	764.5	538.4	774.5	14
Cystatin C	P01034	LVGGPMDASVEEEGVR	y9	823.9	975.4	828.9	985.4	26
Kallikrein-6	Q92876	LSELIQPLPLER	y6	704.3	724.4	709.3	734.4	25
Leucine-rich alpha-2-glycoprotein	P02750	DLLLPQPDLR	y6	590.4	725.4	595.4	735.4	18
Prostaglandin D2 synthase	P41222	AQGFTEDTIVFLPQTDK	Interference in all transitions	N/A	N/A	N/A	N/A	N/A
		WFSAGLASNSSQLR	Interference in all transitions	N/A	N/A	N/A	N/A	N/A
Secretogranin 2	P13521	IILEALR	y5	415	601	420	611	38
Semaphorin-7A	O75326	VYLFDFPEGK	y8	607.7	952.4	611.7	960.4	20
Serotransferrin	P02787	YLGEEYVK	y6	501	724.3	505	732.3	16
Serum albumin	P02768	LVNEVTEFAK	y8	575.8	937.5	579.8	945.5	34
Transthyretin	P02766	AADDTWEPFASGK	y8	697.7	921.4	701.7	929.4	45
Haptoglobin	P00738	TEGDGVYTLNNEK	y7	720.3	881.4	724.3	889.4	29

6.8 Biomarker candidates with literature references

Protein	Accession number	Reference
Alpha-1-acid glycoprotein	P02763	(Rithidech, Honikel et al. 2009)
Alpha-2-macroglobulin	P01023	(Stoop, Dekker et al. 2008) (Ottervald, Franzen et al. 2010)
Ceruloplasmin	P00450	(Stoop, Dekker et al. 2008) (Ottervald, Franzen et al. 2010) (Comabella, Fernandez et al. 2010)
Complement C3	P01024	(Stoop, Dekker et al. 2008) (Ottervald, Franzen et al. 2010)
Neurosepin		
Prostaglandin-H2 D-isomerase	P41222	(Stoop, Dekker et al. 2008)
Semaphorin-7A	O75326	(Comabella, Fernandez et al. 2010) (Comabella, Fernandez et al. 2010)
Transthyretin	P02766	(Stoop, Dekker et al. 2008) (Ottervald, Franzen et al. 2010)
Osteopontin	P10451	
Amyloid beta A4 protein	P05067	(Ottervald, Franzen et al. 2010)
Pigment epithelium derived factor	P36955	(Ottervald, Franzen et al. 2010)
Zinc-alpha-2-glycoprotein	P25311	
Cystatin-C	P01034	
Alpha-1-anti trypsin	P01009	(Ottervald, Franzen et al. 2010)
Haptoglobin		(Comabella, Fernandez et al. 2010)
Ectonucleotide pyrophosphatase	Q13822	
Hemopexin	P02790	(Ottervald, Franzen et al. 2010)